

# **Unterstützung von Aspekten kollaborativer Reflexion durch Prompts in Communities of Practice**

D i s s e r t a t i o n

zur Erlangung des Doktorgrades  
der Naturwissenschaften

vorgelegt von  
Oliver Blunk  
aus Bad Segeberg

genehmigt von der Fakultät für Mathematik/Informatik und Maschinenbau  
der Technischen Universität Clausthal,

Tag der mündlichen Prüfung  
16.01.2020

Vorsitzender der Promotionskommission  
Prof. Dr. Jörg P. Müller

Betreuer  
Prof. Dr.-Ing. Michael Prilla

Gutachter  
Prof. Dr.-Ing. Michael Prilla  
Prof. Dr.-Ing. Thomas Herrmann









## Inhaltsverzeichnis

Inhaltsverzeichnis .....	I
Verzeichnis der Abkürzungen und Akronyme .....	VII
Abbildungsverzeichnis .....	VIII
Tabellenverzeichnis .....	X
1 Einleitung.....	1
1.1 Motivation .....	1
1.2 Forschungsfrage .....	2
1.3 Forschungsansatz.....	4
1.4 Aufbau der Arbeit.....	6
2 Reflexion – Potenziale, Einsatzbereiche und Unterstützungsbedarfe .....	9
2.1 Communities of Practice als Ort für Erfahrungsaustausch und Reflexion.....	9
2.1.1 Was sind Communities of Practice .....	9
2.1.2 Informelles Lernen und Reflexion in Communities of Practice .....	12
2.1.3 Benötigte IT-Unterstützung .....	13
2.2 Reflexion .....	13
2.3 Einsatzbereiche.....	14
2.4 Individuelle Reflexion.....	16
2.4.1 Reflexives Lernen nach Dewey .....	16
2.4.2 Reflexion nach Schön .....	18
2.4.3 Experiential Learning nach Kolb .....	22
2.4.4 Reflexives Lernen nach Boud et al. ....	27
2.4.5 Weitere Aspekte von Reflexion und aktuellere Definitionen .....	32
2.4.6 Ebenen von Reflexion .....	34
2.5 Kollaborative Reflexion .....	35
2.5.1 Begriffsklärung .....	36
2.5.2 Elemente kollaborativer Reflexion .....	38
2.5.3 Gelegenheiten für kollaborative Reflexion .....	40
2.5.4 Der Arbeitsplatz als Reflexionsumgebung .....	41
2.6 Einsatz von Reflexion aus der Sicht eines Unternehmens .....	43
2.7 Abgrenzung von anderen Konzepten .....	44
2.7.1 Productive Reflection.....	44
2.7.2 Critical Reflection .....	44
2.7.3 Critical Thinking .....	45
2.7.4 Reflexivity und Team Reflexivity.....	46
2.7.5 Sensemaking .....	47
2.8 Unterstützungsbedarf und Unterstützungsmöglichkeiten von (kollaborativer) Reflexion .....	47
2.8.1 Herausforderungen in der täglichen Praxis .....	48

2.8.2	Unterstützungsmöglichkeiten (kollaborativer) Reflexion.....	50
2.8.3	Journale .....	51
2.8.4	E-Portfolios .....	52
2.8.5	Visualisierungen.....	53
2.8.6	Prompts .....	54
2.9	Gestaltung von Reflexionsunterstützung.....	54
2.9.1	Modell zur Unterstützung des Reflexionsprozesses .....	54
2.9.2	Gestaltung .....	56
2.10	Prompts als skalierbare und personalisierbare Unterstützung für Reflexion ....	59
2.10.1	Definition .....	59
2.10.2	Einsatz und Wirkung von Prompts .....	61
2.10.3	Design von Prompts .....	66
2.10.4	Arten von Prompts .....	68
2.10.5	Abgrenzung von anderen Begriffen.....	73
2.10.5.1	Fragen .....	73
2.10.5.2	Nudging.....	74
2.10.5.3	Notifikation.....	75
2.11	Zusammenfassung .....	76
3	Konzept zur Unterstützung von Reflexionszyklen durch Prompts.....	79
3.1	Motivation und Ansatz .....	79
3.2	Prompts als Unterstützungsmöglichkeiten für Aspekte kollaborativer Reflexion .....	81
3.3	Studien.....	88
3.4	Qualitative Inhaltsanalyse zur Überprüfung der Prompts .....	90
3.4.1	Kodierschema für Reflexion .....	90
3.4.2	Erweiterung des Kodierschemas für Prompts .....	93
3.4.3	Anwendung des Kodierschemas .....	93
3.4.4	Vorgehen bei der Inhaltsanalyse .....	94
3.5	Aufbau des Softwaresystems als Grundlage für Studie 1 .....	95
3.5.1	Übergeordnete Architektur .....	96
3.5.2	Prompt-Service Backend.....	96
3.5.2.1	Komponenten.....	97
3.5.2.2	Genutzte Frameworks .....	98
3.5.2.3	Webservice-Schnittstelle.....	99
3.5.3	Community of Practice Plattform auf Basis von WordPress.....	101
3.5.3.1	Prompting-Plug-in.....	102
3.5.3.2	Plug-in für nutzergenerierte Prompts.....	107
3.6	Crowdsourcing als Umgebung für Studie 2 .....	107
3.6.1	Einführung .....	107
3.6.2	Einsatzbereiche von Crowdsourcing.....	110
3.6.3	Definition .....	110



3.6.4	Plattformen für Crowdsourcing .....	112
3.6.5	Demografie der Crowdworker .....	113
3.6.6	Qualität der Ergebnisse .....	113
3.6.7	Umgang mit Nutzern mit Täuschungsabsicht .....	114
3.6.8	Nachteile von Crowdsourcing .....	116
3.6.9	Anforderungen an Studiendesign .....	117
4	Studie 1: Prompts als Reflexionsunterstützung in der Praxis öffentlicher Verwaltungen .....	119
4.1	Kontext der Studie .....	119
4.2	Projektkontext EmployID .....	120
4.2.1	Hintergrund und Zielstellung des Forschungsprojekts .....	120
4.2.2	Verankerung der Studie im Projekt .....	122
4.3	Methode .....	123
4.3.1	Studiendesign .....	123
4.3.2	Verwendete Prompts .....	124
4.3.3	Vorgehen für die Analyse .....	128
4.4	Anwendungsfall: Slowenien .....	129
4.4.1	Kontext des Anwendungsfalles in Slowenien .....	129
4.4.2	Deskriptive Analyse des Anwendungsfalles in Slowenien .....	131
4.4.3	Quantitative Analyse des Anwendungsfalles in Slowenien .....	132
4.4.4	Qualitative Analyse des Anwendungsfalles in Slowenien .....	141
4.4.5	Analyse der Interviews mit Nutzern der Community of Practice Plattform .....	143
4.4.6	Diskussion des Anwendungsfalles in Slowenien .....	144
4.5	Anwendungsfall: Kroatien .....	146
4.5.1	Kontext des Anwendungsfalles in Kroatien .....	147
4.5.2	Deskriptive Analyse des Anwendungsfalles in Kroatien .....	148
4.5.3	Quantitative Analyse des Anwendungsfalles in Kroatien .....	149
4.5.4	Qualitative Analyse des Anwendungsfalles in Kroatien .....	158
4.5.5	Diskussion des Anwendungsfalles in Kroatien .....	163
4.6	Allgemeine Diskussion der Anwendungsfälle .....	164
4.7	Einschränkungen .....	167
4.8	Zusammenfassung .....	169
5	Studie 2: Anregung von Reflexionselementen durch Prompts .....	171
5.1	Teilstudie 2.1: Prompts für Erfahrungen und Lösungen im Vergleich .....	171
5.1.1	Ansatz .....	171
5.1.2	Methode .....	172
5.1.2.1	Umsetzung der Teilstudie .....	172
5.1.2.2	Abbildung in Crowdfunder .....	173
5.1.2.3	Probelauf .....	176
5.1.2.4	Methodik für die Analyse .....	176

5.1.3	Hypothesen.....	177
5.1.4	Deskriptive Analyse der Daten .....	178
5.1.5	Qualitative Analyse.....	186
5.1.6	Diskussion.....	191
5.1.6.1	Überprüfung der Hypothesen.....	191
5.1.6.2	Interpretation der übrigen Variablen.....	192
5.1.6.3	Möglichkeiten für das Design von Prompts.....	194
5.1.7	Einschränkungen .....	194
5.1.8	Zusammenfassung.....	196
5.2	Teilstudie 2.2: Spezifität von Prompts im Bereich von Lösungs-Prompts .....	196
5.2.1	Ansatz.....	196
5.2.2	Methode .....	197
5.2.2.1	Umsetzung der Teilstudie .....	197
5.2.2.2	Abbildung in Crowdfunder .....	198
5.2.2.3	Geplante Methodik für die Analyse.....	200
5.2.3	Hypothesen.....	201
5.2.4	Deskriptive Analyse der Daten .....	203
5.2.5	Qualitative Analyse.....	210
5.2.6	Diskussion.....	212
5.2.6.1	Überprüfung der Hypothesen.....	212
5.2.6.2	Interpretation der übrigen Variablen.....	214
5.2.6.3	Möglichkeiten für das Design von Prompts.....	215
5.2.7	Einschränkungen .....	215
5.2.8	Zusammenfassung.....	216
5.3	Teilstudie 2.3: Die Qualität von Lösungsvorschlägen in kollaborativer Reflexion .....	217
5.3.1	Ansatz.....	217
5.3.2	Methodik .....	219
5.3.2.1	Studiendesign.....	219
5.3.2.2	Abbildung in FigureEight .....	220
5.3.2.3	Auswahl und Vorbereitung der Daten .....	221
5.3.2.4	Maße zur Bewertung von Lösungsvorschlägen.....	224
5.3.2.5	Anwendung der Bewertungskriterien .....	229
5.3.2.6	Qualitätssicherung in der Crowdsourcing-Umgebung.....	231
5.3.2.7	Testlauf .....	233
5.3.3	Bereinigung der Studiendaten vor der Analyse .....	234
5.3.4	Demografische Daten der Probanden .....	235
5.3.4.1	Demografische Analyse der Bewertenden.....	236
5.3.4.2	Demografische Analyse der Autoren hinsichtlich der Lösungsvorschläge.....	243

5.3.4.3	Unterschiede in den Probandengruppen zwischen Teilstudie 2.1 und 2.2.....	245
5.3.5	Analyse der Daten.....	246
5.3.5.1	Analyse der Lösungsvorschläge von Teilstudie 2.1 und 2.2	246
5.3.5.2	Regressionsmodelle zur Überprüfung des Einflusses demografischer Variablen.....	250
5.3.5.3	Zusammenhänge zwischen Bewertungskriterien und reflexionsbezogenen Codes .....	267
5.3.5.4	Analyse der Prompts .....	269
5.3.5.5	Weitere Analysen.....	270
5.3.6	Diskussion.....	272
5.3.6.1	Überprüfung der Kriterien .....	272
5.3.6.2	Diskussion der übrigen Ergebnisse.....	273
5.3.6.3	Implikationen für das Design von Reflexionsunterstützung.....	276
5.3.7	Einschränkungen.....	277
5.3.8	Zusammenfassung.....	279
5.4	Zusammenfassende Betrachtung der Forschungsfragen .....	279
6	Diskussion der Ergebnisse beider Studien.....	281
6.1	Diskussion der Ergebnisse.....	281
6.1.1	Häufigkeit von Beiträgen .....	281
6.1.2	Nutzergenerierte Prompts .....	282
6.1.3	Effekte von Prompts.....	283
6.1.4	Zeitpunkt von Lösungs-Prompts in der Reflexion.....	286
6.1.5	Anregung von neuartigen oder ungewöhnlichen Ideen .....	288
6.1.6	Unerwartete Erlebnisse als Gegenstand in Prompts .....	289
6.1.7	Anzahl der Wörter als Faktor.....	290
6.1.8	Mögliche Einflüsse der demografischen Daten .....	290
6.1.9	Diskussion des Layouts auf der Community-Plattform.....	292
6.1.9.1	Design der Prompts.....	292
6.1.9.2	Startseite der Community-Plattform.....	292
6.2	Diskussion der Ergebnisse der Studien hinsichtlich der Forschungsfrage.....	293
6.3	Einschränkungen der Arbeit.....	294
7	Empfehlungen zur Implementierung von Prompting als Unterstützung von Reflexion .....	297
7.1	Empfehlungen auf Basis konkreter Ergebnisse.....	297
7.2	Vorschläge für Prompt-Design.....	298
7.2.1	Vorschlag 1: Hervorhebung von nutzergenerierten Prompts.....	299
7.2.2	Vorschlag 2: Positionierung von Prompts .....	300
7.2.3	Vorschlag 3: Prompts zur Vermeidung sehr kurzer Beiträge .....	302
7.2.4	Vorschlag 4: Konvergenzphase zur Verarbeitung der Lösungsvorschläge.....	303

8 Zusammenfassung .....	305
8.1 Möglichkeiten für weitere Forschung .....	305
8.2 Fazit .....	306
9 Anhang .....	311
A Vorgehen und Grundlagen für die statistische Analyse in Teilstudie 2.1 .....	311
B Vorgehen und Grundlagen für die statistische Analyse in Teilstudie 2.2 .....	313
C Vorgehen und Grundlagen für die statistische Analyse in Teilstudie 2.3 .....	316
D Statistiken für Teilstudie 2.3 – Demografische Daten .....	325
E Ordinale und Multinominale Regressionsmodelle für Teilstudie 2.3 .....	328
Literaturverzeichnis .....	345

## Verzeichnis der Abkürzungen und Akronyme

AGR	Inhaltsanalyse: Zustimmung
ADV	Inhaltsanalyse: Hinweis
AMT	Amazon Mechanical Turk
CHANGE	Inhaltsanalyse: Wandel
CMS	Content-Management-System
CoP	Community of Practice
DISAGR	Inhaltsanalyse: Ablehnung
D_LOOP	Inhaltsanalyse: Double Loop Learning
EMO_OWN	Inhaltsanalyse: Eigene Emotionen
EMO_OTH	Inhaltsanalyse: Emotionen Dritter
EP	Erfahrungs-Prompt
EXP	Inhaltsanalyse: Erfahrung
HIT	Human Intelligence Task
HZZ	Hrvatski zavod za zapošljavanje (kroatische Arbeitsagentur)
INT	Inhaltsanalyse: Interpretation
KLP	Spezifischer Kontroll-Prompt
KNO	Inhaltsanalyse: Wissen
KP	Kontroll-Prompt
LDAP	Lightweight Directory Access Protocol
LP	Lösungs-Prompt
Q_INF	Inhaltsanalyse: Fragen nach Information
Q_INT	Inhaltsanalyse: Fragen nach Interpretation
SLP	Spezifischer Lösungs-Prompt
St. Abw.	Standardabweichung
SUG	Inhaltsanalyse: Lösungsvorschlag
SUG_EXP	Inhaltsanalyse: Erfahrungsbasierter Lösungsvorschlag
SUG_KNO	Inhaltsanalyse: Wissensbasierter Lösungsvorschlag
S_LOOP	Inhaltsanalyse: Single Loop Learning
YEP	Youth Employment Project
ZRSZ	Zavod Republike Slovenije za zaposlovanje (slowenische Arbeitsagentur)

## Abbildungsverzeichnis

Abbildung 2-1: Experiential learning Zyklus nach Kolb und Fry (Kolb und Fry 1975, Figure 1).....	23
Abbildung 2-2: Reflexives Lernen nach Boud et al. (1985, 36, Figure 3) .....	29
Abbildung 2-3: Kollaborative Reflexion und die Rolle von Artikulation (Prilla, Degeling und Herrmann 2012, Figure 3; Prilla und Renner 2014, Figure 1) .....	39
Abbildung 2-4: Computer-supported reflective learning model nach Krogstie et al. (2013, Fig. 1) .....	55
Abbildung 2-5: Unterscheidung zwischen dem Sender, der über einen Prompting-Mechanismus einen auditiven, visuellen oder textuellen Prompt an den Empfänger schickt .....	60
Abbildung 3-1: Übersicht der Architektur.....	96
Abbildung 3-2: Prompts auf der Startseite der Plattform (unten, gelb hinterlegt) .....	101
Abbildung 3-3: Prompts im Bereich der Foren (grün hinterlegt).....	104
Abbildung 4-1: Theoriehintergrund in EmployID (Bimrose u. a. 2014a).....	121
Abbildung 4-2: Prompts im Bereich der Foren (grün hinterlegt).....	126
Abbildung 4-3: Prompts auf der Startseite der Plattform (unten, gelb hinterlegt) .....	127
Abbildung 5-1: Simulierte Forenumgebung in Crowdfower mit einem Thema, zu dem eine Antwort verfasst werden soll. Es wird im Bild des Kontroll-Prompts angezeigt. ....	174
Abbildung 5-2: Eigene Auskunft zur wöchentlichen Arbeitszeit als Crowdwoker .....	182
Abbildung 5-3: Eigene Auskunft zur Arbeitserfahrung als Crowdwoker in Teilstudie 2.1 .....	183
Abbildung 5-4: Eigene Auskunft zum Bildungsgrad der Probanden in Teilstudie 2.1 .....	185
Abbildung 5-5: Angabe der Länder aus denen die Crowdwoker aus Teilstudie 2.1. Angaben von Crowdfower. ....	185
Abbildung 5-6: Simulierte Diskussion in Teilstudie 2.1 in Crowdfower.....	200
Abbildung 5-7: Verteilung der wöchentlichen Arbeitsstunden der Crowdwoker in Teilstudie 2.2 aufgeteilt nach Prompts .....	206
Abbildung 5-8: Erfahrung der Crowdwoker in Teilstudie 2.2 aufgeteilt nach Prompts .....	207
Abbildung 5-9: Bildungsgrad der Crowdwoker in Teilstudie 2.2 je Prompt.....	209
Abbildung 5-10: Land der Crowdwoker in Teilstudie 2.2 laut Crowdfower aufgeteilt nach Prompts .....	209
Abbildung 5-11: Abbildung von Teilstudie 2.3 in FigureEight .....	220
Abbildung 5-12: Wöchentliche Arbeitsstunden der Probanden in Teilstudie 2.3 (n=134) .....	239

Abbildung 5-13: Erfahrung der Probanden mit Crowdsourcing in Teilstudie 2.3 (n=134)	240
Abbildung 5-14: Bildungsgrad der Probanden in Teilstudie 2.3 (n=134)	241
Abbildung 5-15: Land der Probanden in Teilstudie 2.3 (n=134)	242
Abbildung 7-1: Prompts im Bereich der Foren (grün hinterlegt; ursprüngliches Design)	299
Abbildung 7-2: Prototyp von verbesserten nutzergenerierten Prompts (eigener Entwurf; neues Design)	300
Abbildung 7-3: Entwurf einer Prompt-Wiederholung (eigener Entwurf)	301
Abbildung 7-4: Entwurf eines Hinweises für die Anzahl der Wörter (eigener Entwurf)	303

## Tabellenverzeichnis

Tabelle 2-1: Unterschiedliche Arten von Prompts .....	68
Tabelle 3-1: Liste von Prompts zu den jeweiligen Zielen .....	86
Tabelle 3-2: Reflexionsbezogenes Kodierschema (angepasst aus (Prilla u. a. 2015); Beschreibungen wurden übersetzt und Beispiele sind übernommen.) .....	91
Tabelle 3-3: Verwendete Frameworks in der Software .....	98
Tabelle 3-4: Übersicht welche Funktionen die Schnittstelle zum Abrufen von Prompts bereitstellt. Für jede Funktion ist auch der Pfad für den Aufruf angegeben, und um welche Art von Aufruf es sich handelt. ....	99
Tabelle 4-1: Übersicht der verwendeten Prompts, wo (innerhalb einer Diskussion, neue Diskussion) diese sichtbar waren, und welche Rolle (Autor, Leser, Antwortender) im Forum den Prompt empfangen konnte und welches Ziel die Prompts aus dem Konzept abdecken (Abschnitt 3.2) .....	125
Tabelle 4-2: Ablauf der Studie in Slowenien .....	130
Tabelle 4-3: Aktivität der Benutzer und Moderatoren in der Community in Slowenien zwischen 02.2016 und 02.2018 ohne Gruppen für E-Coaching oder DrogArt .....	131
Tabelle 4-4: Ursprünglicher Zeitraum, ohne die ausgeschlossenen Nutzer (anonymisierte Nutzer-IDs u/4996 und u/2798), ohne die Aktivitäten in den Gruppen für Peer-Coaching-Kurs und für schwervermittelbare Arbeitssuchende (DrogArt) .....	133
Tabelle 4-5: Kompletter Zeitraum, ohne die ausgeschlossenen Nutzer (anonymisierte Nutzer-IDs u/4996 und u/2798), ohne Peer-Coaching-Kurs, ohne DrogArt .....	133
Tabelle 4-6: Übersicht welcher Prompt in Slowenien wie oft gesehen wurde und wie oft ein Beitrag geschrieben wurde (sortiert nach Ort).....	137
Tabelle 4-7: Sichtbarkeit pro Prompt-Kategorie und Ort in Slowenien.....	140
Tabelle 4-8: Häufigkeit des Einflusses eines Prompts auf ein Thema (Slowenien).....	142
Tabelle 4-9: Allgemeine Korrelation Einfluss und die reflexionsbezogenen Codes (n=277 Beiträge, bei denen in 24 Beiträgen ein Einfluss kodiert wurde; 66x Code SUG_EXP kodiert).....	142
Tabelle 4-10: Einfluss von nutzergenerierten Prompts (n=30 Beiträge, bei denen in 19 Beiträgen ein Einfluss kodiert wurde; 8x Code SUG_EXP und 13x Code Q_INT kodiert) .....	143
Tabelle 4-11: Korrelation vom lernbezogenen Prompt (n=33 Beiträge, bei denen in 2 Beiträgen ein Einfluss kodiert wurde; 1x Code D_LOOP und 2x Code Q_INF kodiert) .....	143
Tabelle 4-12: Ablauf der Studie in Kroatien .....	147
Tabelle 4-13: Gesamtaktivität der Benutzer und Moderatoren in der Community in Kroatien zwischen 10.2016 und 02.2018 .....	148



Tabelle 4-14: Lese- und Schreibraten der Nutzer (ohne Nutzer mit anonymisierter ID u/18137) und Moderatoren in den beiden Konditionen der Studie (nur Zeiträume der Konditionen) .....	149
Tabelle 4-15: Lese- und Schreibraten der Nutzer (Nutzer ohne mit anonymisierter ID u/18137) und Moderatoren in den beiden Konditionen der Studie (kompletter Datenerfassungszeitraum) .....	151
Tabelle 4-16: Übersicht welcher Prompt in Kroatien wie oft gesehen wurde und wie oft ein Beitrag geschrieben wurde (gruppiert nach den Orten: Prompts bei Erstellung neuer Themen, bei der Erstellung neuer Antworten und auf der Startseite (siehe auch Abbildung 4-2 und Abbildung 4-3 auf S. 127)) .....	154
Tabelle 4-17: Sichtbarkeit pro Prompt-Kategorie und Ort in Kroatien .....	156
Tabelle 4-18: Häufigkeit des Einflusses eines Prompts auf ein Thema (Kroatien) .....	158
Tabelle 4-19: Korrelation zwischen den Prompts mit Einfluss auf den Text und den Codes für Reflexion (n=203 Beiträge, bei denen in 32 Beiträgen ein Einfluss kodiert wurde; 107x Code EXP, 12x Code ADV und 67x Code SUG_EXP kodiert) .....	159
Tabelle 4-20: Korrelationen beim Einfluss von nutzergenerierten Prompts auf den Text (n=21 Beiträge, von denen in 11 Beiträgen ein Einfluss kodiert wurde; 13x Code SUG_EXP und 5x Code Q_INT) .....	160
Tabelle 4-21: Verbindung zwischen dem lernbezogenen Prompt und den reflexionsbezogenen Codes in Kroatien (n=42, von denen in einem Beitrag ein Einfluss kodiert wurde; 2x Code DISAGR und 4x Code AGR) .....	161
Tabelle 4-22: Übersicht wie oft bei einem Prompt-Einfluss auf den Text kodiert wurde .....	162
Tabelle 5-1: Prompts in Teilstudie 2.1 .....	173
Tabelle 5-2: Übersicht der deskriptiven Daten in Teilstudie 2.1 .....	179
Tabelle 5-3: Anzahl der Werte für jede Gruppe in der gruppierten Variable für die Zeit .....	180
Tabelle 5-4: Anteil wie oft der Code einem Text je nach Anzahl der Wörter hinzugefügt wurde .....	180
Tabelle 5-5: Einfluss von Prompts und der Anzahl der Wörter auf Erfahrungen (Code EXP) .....	187
Tabelle 5-6: Einfluss der Prompts und der Anzahl der Wörter auf allgemeine Lösungsvorschläge (Code SUG) .....	188
Tabelle 5-7: Einfluss der Prompts, der Sprache und der Anzahl der Wörter auf erfahrungsbasierte Lösungsvorschläge (Code SUG_EXP) .....	189
Tabelle 5-8: Einfluss der Prompts und der Sprache auf Hinweise (Code ADV) .....	190
Tabelle 5-9: Einfluss der Prompts auf wissensbasierte Lösungsvorschläge (Code SUG_KNO) .....	190
Tabelle 5-10: Prompts in Teilstudie 2.2 .....	198

Tabelle 5-11: Neue Codes für die inhaltliche Analyse. Ein Prompt schlägt die Nutzung von Vorschlägen aus den vordefinierten Antworten vor. In der Spalte „Kommentar“ ist abgebildet, welcher Vorschlag, in welchem Kommentar unterbreitet wird.....	201
Tabelle 5-12: Deskriptive Daten in Teilstudie 2.1 (alle Personen für Vergleich innerhalb der Teilstudie).....	204
Tabelle 5-13: Anzahl der Werte für jede Gruppe in der gruppierten Variable für die Zeit .....	205
Tabelle 5-14: Anteil, wie oft ein Prompt je nach Anzahl der Wörter einem Text hinzugefügt wurde .....	205
Tabelle 5-15: Einflüsse der Variablen NativeSpeaker und Anzahl der Wörter auf den Code SUG.....	211
Tabelle 5-16: Einflüsse der Variablen NativeSpeaker und Anzahl der Wörter auf SUG_EXP.....	212
Tabelle 5-17: Übersicht über die Bearbeitungsschritte in der Datenaufbereitung .....	224
Tabelle 5-18: Spanne von Bewertungskriterien in der Literatur .....	229
Tabelle 5-19: Bewertung des Probelaufs in Teilstudie 2.3 (n=5).....	233
Tabelle 5-20: Abschließende Beurteilung des Jobs durch die Crowdworker (n=30 von 134 Personen im Job).....	236
Tabelle 5-21: Demografische Informationen über die Probanden in Teilstudie 2.3 (n=134) .....	237
Tabelle 5-22: Verteilung der Bewertungen innerhalb der Kategorien gruppiert nach Altersgruppe (3er-Likert-Skala) .....	238
Tabelle 5-23: Verteilung der Bewertung ungewöhnlicher Ideen für Männer und Frauen (5er-Likert-Skala) .....	239
Tabelle 5-24: Signifikante Verteilungen der Bewertungskriterien innerhalb der Variable Erfahrung mehr oder weniger als ein Jahr (5er-Likert-Skala; mit * markierte Zeilen, zeigen signifikante Zusammenhänge bei der berechneten 3er Likert-Skala) .....	240
Tabelle 5-25: Zusammenhänge zwischen den Bewertungskriterien und der Variable <i>Academics</i> (3er-Likert-Skala).....	242
Tabelle 5-26: Verteilung der Bewertung von Personen mit Englisch als Muttersprache (3er-Likert-Skala) .....	243
Tabelle 5-27: Zusammenhang zwischen der Variable ExpMoreThanAYear der Autoren und dem Kriterium für die Klarheit der Formulierung.....	244
Tabelle 5-28: Zusammenhang zwischen der Variable <i>Academics</i> der Autoren und den Bewertungskriterien.....	244
Tabelle 5-29: Zusammenhang zwischen der Variable NativeSpeaker der Autoren und den Bewertungskriterien.....	245
Tabelle 5-30: Mann-Whitney-U Tests des Unterschiedes zwischen den Probandengruppen in den beiden Teilstudien 2.1 und 2.2. Es wurden nur	

Beiträge, die bei dem Lösungs-Prompt entstanden sind, berücksichtigt. Signifikante Unterschiede sind mit * markiert. ....	246
Tabelle 5-31: Durchschnitt und Standardabweichung der Bewertungen je Kategorie und Studie .....	247
Tabelle 5-32: Mann-Whitney-U Test der (*signifikanten) Bewertungskategorien zwischen den Konditionen mit und ohne Diskussionen (alle Prompts) .....	248
Tabelle 5-33: Mann-Whitney-U Test der (*signifikanten) Bewertungskategorien für Vorschläge aus Lösungs-Prompts zwischen den Konditionen mit (M) und ohne (O) Diskussionen (n=331) .....	249
Tabelle 5-34: Vergleich der Bewertungen für den Lösungs-Prompt zwischen Teilstudie 2.1 und 2.2 (5er-Likert-Skala, außer 3er-Likert-Skala bei K12) .....	250
Tabelle 5-35: Ordinales Regressionsmodell für das Kriterium für ungewöhnliche Elemente in einem Lösungsvorschlag (K2). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird. ....	252
Tabelle 5-36: Ordinales Regressionsmodell für das Kriterium den Grad wie sehr der Lösungsvorschlag das Problem lösen kann (K5). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird. ....	254
Tabelle 5-37: Ordinales Regressionsmodell für das Kriterium den Grad der Glaubwürdigkeit (K7). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird. ....	256
Tabelle 5-38: Ordinales Regressionsmodell für das Kriterium den Grad der Klarheit der Beschreibung (K8). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird. ....	258
Tabelle 5-39: Ordinales Regressionsmodell für das Kriterium den Grad wie leicht der Lösungsvorschlag in die Praxis überführbar ist (K9). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird. ....	260
Tabelle 5-40: Ordinales Regressionsmodell für das Kriterium den Grad wie sehr der Lösungsvorschlag eine Auswirkung auf die künftigen Handlungen des Bewertenden hat (K10). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird. ....	262
Tabelle 5-41: Ordinales Regressionsmodell für das Kriterium den Grad wie sehr der Bewertende den Lösungsvorschlag weiterempfehlen würde (K11). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird. ....	264

Tabelle 5-42: Ordinales Regressionsmodell für das Kriterium den Grad wie sehr der Bewertende den Lösungsvorschlag nutzen würde, sofern er das Problem hätte (K12). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird. ....	265
Tabelle 5-43: Verteilungen der Bewertungen innerhalb der einzelnen reflexionsbezogenen Codes aufgeteilt nach Bewertungskategorie (5er-Likert-Skala). ....	268
Tabelle 5-44: Mann-Whitney-U-Test zum Vergleich der Qualitäten der erfahrungsbasierten Lösungsvorschläge aus Teilstudie 2.1 generiert aus Lösungs-Prompts und Erfahrungs-Prompts (n=300). ....	269
Tabelle 5-45: Vergleich der Bewertungen für den Lösungs- und Erfahrungs-Prompt in Teilstudie 2.1, bei denen sich die Bewertungen signifikant unterscheiden (5er-Likert-Skala) .....	270
Tabelle 5-46: Mann-Whitney-U-Test zum Vergleich der Qualitäten der erfahrungsbasierten Lösungsvorschläge aus Teilstudie 2.2, berechnet aus Kontroll-Lösungs-Prompt und Lösungs-Prompt (n=384) .....	270
Tabelle 9-1: Value inflated factors zur Überprüfung der Multikollinearität in Teilstudie 2.1 .....	312
Tabelle 9-2: Eigenvalues zur Überprüfung der Multikollinearität in Teilstudie 2.1 .....	313
Tabelle 9-3: Value inflated factors zur Überprüfung der Multikollinearität in Teilstudie 2.2 .....	315
Tabelle 9-4: Überprüfung der Multikollinearität in Teilstudie 2.1 .....	316
Tabelle 9-5: Übersicht der Eigenschaften der ordinalen Regressionsmodelle (5er-Likert-Skala), die aus allen demografischen Daten (sowohl der Autoren der Beiträge als auch der Bewertenden) sowie einer Variable für die Diskussionsbeiträge (Unterscheidung Teilstudien 2.1 und 2.2) bestehen. Es wurden Regressionsmodelle für alle Kategorien und für alle Texte berechnet, die bei einem Lösungs-Prompt entstanden sind. Valide Regressionsmodelle sind mit * markiert.....	321
Tabelle 9-6: Überprüfung der Multikollinearität in Teilstudie 2.3. In der Tabelle werden alle demografischen Daten sowohl der Autoren (erhoben in Teilstudie 2.2) als auch der Bewertenden (erhoben in Teilstudie 2.3) überprüft.....	322
Tabelle 9-7: Übersicht der Eigenschaften der ordinalen Regressionsmodelle (3er-Likert-Skala), die aus allen demografischen Daten (sowohl der Autoren der Beiträge als auch der Bewertenden) sowie einer Variable für die Diskussionsbeiträge (Unterscheidung Teilstudien 2.1 und 2.2) bestehen. Es wurden Regressionsmodelle für alle Kategorien und für alle Texte berechnet, die bei einem Lösungs-Prompt entstanden sind. Valide Regressionsmodelle sind mit * markiert.....	323
Tabelle 9-8: Test auf Normalverteilung der Bewertungskategorien. ....	324
Tabelle 9-9: Zusammenhang zwischen Bewertungskriterien und der Variable Academics.....	325

Tabelle 9-10: Zusammenhang je Bewertung und der Anzahl der Wörter (3er-Likert-Skala) .....	326
Tabelle 9-11: Zusammenhang zwischen der Variable für die Wochenstunden der Autoren und den Bewertungskriterien .....	327
Tabelle 9-12: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie ungewöhnlich der Lösungsvorschlag ist (K2). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-7 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet. ....	328
Tabelle 9-13: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie sehr der Lösungsvorschlag das Problem lösen kann (K5). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 79 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet. ....	330
Tabelle 9-14: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie glaubwürdig der Lösungsvorschlag ist (K7). Der Test für die Validität des Regressionsmodells auf Basis der 5er-Likert-Skala ist in Tabelle 9-5 abgebildet. Der Referenzwert für den Vergleich ist die Bewertung <i>strongly disagree</i> . Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet. ....	332
Tabelle 9-15: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie klar der Lösungsvorschlag formuliert ist (K8). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-7 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet. ....	334
Tabelle 9-16: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie einfach der Lösungsvorschlag in die Praxis überführt werden kann (K9). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-7 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet. ....	336
Tabelle 9-17: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie Einfluss der Lösungsvorschlag auf künftige Handlungen des Bewertenden nehmen kann (K10). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-7 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet. ....	338
Tabelle 9-18: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich zu welchem Grad der Bewertende den Lösungsvorschlag weiterempfehlen würde (K11). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-5 abgebildet. Der Referenzwert für den	

Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet. ....	340
Tabelle 9-19: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich zu welchem Grad der Bewertende den Lösungsvorschlag selbst nutzen würde (K12). Der Test für die Validität des Regressionsmodells auf Basis der 3er- Likert-Skala ist in Tabelle 9-5 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet. ....	
	342

# 1 Einleitung

Dieses Kapitel beschreibt die Motivation, der Zielrichtung und den Aufbau der vorliegenden Arbeit.

## 1.1 Motivation

Als informelles Lernen wird Lernen bezeichnet, welches außerhalb formeller Rahmen wie Trainings oder Schulungen stattfindet (Eraut 2004). Dies ist eine wichtige Form des Lernens und viele Studien vermuten, dass bei Berufstätigen bis zu 70 % der Zeit für Lernen im Bereich informellen Lernens stattfindet (Clardy 2018). Ein Mittel informellen Lernens im beruflichen Kontext ist Reflexion (Eraut 2004). Als Reflexion wird das Zurückkehren zu vergangenen Erfahrungen, deren Neubewertung und die daraus resultierende Ableitung von Ideen für Handlungsweisen für die Zukunft bezeichnet (Dewey 1933; Schön 1983; Kolb 1984; Boud, Keogh und Walker 1985). Reflexion wird in der Literatur immer als zielgerichteter Prozess verstanden (Boud, Keogh und Walker 1985), dessen Bestreben das Entwickeln von neuen Herangehensweisen ist, um ähnliche Situationen, wie diejenige, über die reflektiert wurde, in Zukunft besser bestehen zu können. Reflexion ermöglicht es aus den eigenen Erfahrungen zu lernen und wird in den unterschiedlichsten Berufen und auch in der Ausbildung bzw. Studium eingesetzt, wie bspw. bei Gesundheits- und Krankenpflegern (Sandars 2009; Kim u. a. 2010; Platzer, Snelling und Blake 1997; Chirema 2007), Musikern (Johnston, Amitani und Edmonds 2005), Medizinerinnen (Ryan 2010), Psychologen (Hettich 1990), etc.

Reflexion ist nicht nur ein individueller Prozess, sondern kann zusätzlich kollaborativ mit anderen Personen durchgeführt werden, um gegenseitig von den Erfahrungen lernen zu können (Dyke 2006; Prilla und Renner 2014). In der täglichen Praxis kann so etwas bspw. in einer Übergabe zwischen Gesundheits- und Krankenpflegern stattfinden (Prilla, Degeling und Herrmann 2012). In großen Unternehmen sind sogenannte Communities of Practice eine Möglichkeit des Austausches (Wenger, McDermott und Snyder 2002). Unter einer Community of Practice versteht man einen Zusammenschluss von Personen, die eine ähnliche berufliche Praxis ausüben und das gemeinsame Ziel des Lernens haben (Wenger, McDermott und Snyder 2002). Diese Communities of Practice sind ein Ort, an dem sich Personen über ihre Praxis und ihre Erfahrungen austauschen und voneinander lernen. Dementsprechend eignen sich Communities of Practice als Orte für Reflexion als Form informellen Lernens (Vince 2002; Clouder 2000; Raelin 2002; Yang 2009; Stoszkowski und Collins 2017). Communities of Practice können durch geeignete Softwareplattformen unterstützt werden, wenn einzelne Personen oder Teams

geografisch verteilt arbeiten und sich austauschen möchten (Blunk und Prilla 2017a). Dieser Ort kann dann ebenfalls für Reflexion genutzt werden.

In der Praxis hingegen steht Reflexion als Form informellen Lernens unterschiedlichen Herausforderungen gegenüber. Zum einen fehlt in der beruflichen Praxis häufig die Zeit (bspw. Renner u. a. 2016; Hatton und Smith 1995; Moon 1999, 166) und die Gelegenheit (Hatton und Smith 1995), um zu reflektieren. Oft ist aus Sicht eines Unternehmens das tägliche informelle Lernen im Widerspruch zu den übrigen Kerntätigkeiten, mit denen das Unternehmen sein Geschäft durchführt (Blunk und Prilla 2017a). Zudem fehlt bei der Reflexion in der Praxis häufig eine Systematik (Cheetham und Chivers 2000; Renner u. a. 2019). Andere Studien haben gezeigt, dass Reflexion in der Praxis schwierig durchzuführen ist, selbst wenn jemand das Wissen hat, wie Reflexion funktioniert (Platzer, Snelling und Blake 1997).

Um Reflexion in der täglichen Praxis zu fördern, ist Unterstützung notwendig. Ein häufiges Mittel ist reflexives Schreiben, indem bspw. Journale geführt werden (Moon 2006b), oder E-Portfolios, in denen eigene Notizen oder Lerninhalte gesammelt werden (Alexiou und Paraskeva 2010; Elango, Jutti und Lee 2005). Häufige Nachteile an diesen Formen der Unterstützung sind, dass Journale von einem Mentor begleitet werden, wodurch die Form der Unterstützung schwer skaliert. Eine skalierbare Alternative zur Unterstützung von Reflexion bzw. reflexivem Schreiben sind Prompts. Darunter versteht man oft textuelle Hinweise, die dem Empfänger versuchen Hilfestellung anzubieten (Bannert 2009; Thillmann u. a. 2009). Prompts sind eine Art der Unterstützung, die dem Empfänger offenlässt, ob dieser den Hinweis nutzen möchte (Fivush, Gray und Fromhoff 1987). Prompts haben in Studien bereits Reflexion in einem schulischen Umfeld unterstützen können (bspw. Davis und Linn 2000). Dennoch ist unklar, wie ein Prompt eingesetzt werden muss, um für Reflexion eine gute Unterstützung zu bieten. Dies soll in dieser Arbeit untersucht werden und wird im folgenden Abschnitt konkreter beschrieben.

## **1.2 Forschungsfrage**

Der vorherige Abschnitt hat gezeigt, dass eine Community of Practice ein Ort ist, an dem Personen gleicher Praxis informell voneinander lernen können. Reflexion bzw. kollaborative Reflexion ist ein Prozess, bei dem aus Erfahrungen gelernt wird. Dieser unterliegt aber in der Praxis unterschiedlichen Herausforderungen und benötigt daher Unterstützung. Prompts sind eine Form möglicher Unterstützung für (kollaborative) Reflexion am Arbeitsplatz bspw. in Communities of Practice.



Prompts als Form der Unterstützung von Reflexion wurden bereits in mehreren Studien untersucht (bspw. Davis und Linn 2000; Thillmann u. a. 2009; Glogger u. a. 2009; Lai und Calandra 2007; Davis 2003; Berthold, Nückles und Renkl 2007; van den Boom, Paas und van Merriënboer 2007; Papadopoulos u. a. 2009; Clarke 2004). Diese Studien haben gemeinsam, dass sie sich überwiegend auf den schulischen oder universitären Kontext konzentrieren und dort oft individuelle Reflexion im Fokus steht. Der Fokus der Studien ist häufig reflexives Schreiben, wie bspw. in Journalen.

Die genannten Studien fokussieren sich überwiegend auf individuelle Reflexion. Die kollaborative Reflexion mit anderen Personen wird oft vernachlässigt. Dies ist bei anderen Unterstützungsformen für Reflexion ebenfalls der Fall (Renner u. a. 2016). Bei kollaborativer Reflexion steht insbesondere der Austausch von Erfahrungen und das gemeinsame Entwickeln und Diskutieren von Ideen und Lösungsvorschlägen im Vordergrund (Daudelin 1996; Høyrup 2004; Woerkom und Croon 2008; Raelin 2002; Dyke 2006; Hatton und Smith 1995). Dementsprechend sind im Vergleich zur individuellen Reflexion zusätzliche Aktivitäten vorhanden, um das Ziel, dem Lernen aus Erfahrung bzw. dem Erarbeiten neuer Handlungsweisen für zukünftige Situationen, zu erreichen. Hier gibt es wenig Studien, die prüfen, inwieweit sich diese Aktivitäten unterstützen lassen.

Der berufliche Kontext findet in den genannten Studien wesentlich weniger Beachtung. Bei Renner et al. sind bspw. Studien enthalten, die mit Studenten durchgeführt wurden (2016). Andere Studien zeigen jedoch, dass sich Ergebnisse aus Studien mit Probanden aus dem universitären Kontext und Probanden aus einem beruflichen Kontext unterscheiden (Ward 1993). Dementsprechend sind die Ergebnisse aus den Studien nur eingeschränkt auf einen beruflichen Kontext oder den Kontext einer schriftlichen kollaborativen Online-Reflexion zu übertragen. Ebenfalls beinhalten diese Studien keine Aussagen zu Reflexion in Communities of Practice in einem beruflichen Kontext.

Aufgrund der bisherigen Ausrichtung vergangener Studien ist in der Literatur wenig über die Wirkung von Prompts als Unterstützungsfunktion zum Anregen kollaborativer Reflexion bzw. einzelner Aspekte dieser Reflexion in einem beruflichen Kontext bekannt. Dementsprechend lautet die Forschungsfrage dieser Arbeit: *Inwieweit lassen sich Aspekte kollaborativer Reflexion in Communities of Practice in einem beruflichen Kontext durch Prompts anregen?*

Im folgenden Abschnitt wird beschrieben, wie die Forschungsfrage in der vorliegenden Arbeit bearbeitet werden soll.

### 1.3 Forschungsansatz

Wie bisher beschrieben ist (kollaborative) Reflexion ein Mittel informellen Lernens basierend auf den eigenen Erfahrungen (oder den Erfahrungen Dritter). Zunächst soll die Literatur zu (kollaborativer) Reflexion analysiert werden, um die zu unterstützenden Aspekte zu identifizieren. Dazu gehören bspw. das Modell zu individueller Reflexion von Boud et al. (1985) oder spezifischere Modelle zur Unterstützung von kollaborativer Reflexion durch Technologie (Krogstie, Prilla und Pammer 2013).

Die Literatur zu Prompts wird ebenfalls aufgearbeitet. Hier ist die Zielsetzung Hinweise zur Gestaltung von Prompts bzw. der Wirkung unterschiedlicher Prompts zu identifizieren, um diese für die Erstellung eigener Prompts nutzen zu können.

Zusätzlich soll die Literatur hinsichtlich (kollaborativer) Reflexion und Communities of Practice untersucht werden, um mögliche Hinweise zur Erstellung von Unterstützung durch Prompts erarbeiten zu können.

Basierend auf der existierenden Literatur zu individueller und kollaborativer Reflexion und zur Nutzung von Prompts zur Unterstützung von Reflexion wird ein eigenes Konzept erarbeitet. In diesem werden einzelne Aspekte von Reflexion, wie bspw. der Austausch von Erfahrungen in der kollaborativen Reflexion, identifiziert. Zu jedem dieser identifizierten Aspekte von Reflexion werden Prompts erstellt, die diesen Aspekt von Reflexion in der praktischen Durchführung unterstützen sollen. Die vorliegende Arbeit konzentriert sich auf textuelle Prompts (im Vergleich zu Bildern oder auditiven Prompts).

Die Zielsetzung der vorliegenden Arbeit ist die Unterstützung kollaborativer Reflexion in Communities of Practice in einem beruflichen Kontext. Um diese Zielsetzung zu verfolgen, wird das entwickelte Konzept in einer Community of Practice Plattform implementiert und anschließend in einer Feldstudie in zwei Anwendungsfällen evaluiert. Der Vorteil einer Feldstudie ist, dass das Konzept in einer für die Nutzer natürlichen Umgebung implementiert wird. Dies ermöglicht eine aussagekräftigere Analyse hinsichtlich der Prompts als Unterstützung von Reflexion in der Praxis. Zwei Gruppen von Nutzern befinden sich über einen bestimmten Zeitraum in zwei Konditionen (Prompts vs. keine Prompts), um zu evaluieren, inwieweit sich das Vorhandensein von Prompts auf die Antwortrate der Nutzer und die Inhalte der entstandenen Texte auswirkt. Auf diese Weise soll überprüft werden, ob Prompts in der täglichen Praxis als Unterstützung von Reflexion eine Rolle spielen.

Der Nachteil einer Feldstudie, die über einen längeren Zeitraum läuft, ist, dass es viele Einflüsse gibt, die sich nicht kontrollieren lassen, wie bspw. aufgrund von Fluktuation

wechselnde Personen, die diskutieren oder sich ändernde Richtlinien im Unternehmen. Da die Anzahl der Beiträge zur Laufzeit der Feldstudie niedrig ausgefallen ist, wurde parallel eine Reihe von zusätzlichen Studien begonnen.

Die Aspekte Austausch von Erfahrungen und Erarbeiten von Lösungsvorschlägen sollen gezielt in einer Crowdsourcing-Studie evaluiert werden. Innerhalb der Studie wird eine Diskussionsumgebung mit einem für Crowdworker relevanten Problem simuliert. Dies bildet eine Art Arbeitsumgebung für Crowdworker nach, da diese sich gerne in Foren zu ihrer Tätigkeit austauschen (Martin u. a. 2014). Die Probanden sollen in einem neu angelegten Thema eigene Antworten verfassen. Diese Studie soll darüber Aufschluss geben, ob sich die beiden genannten Aspekte von kollaborativer Reflexion anregen lassen und ob es einen Unterschied in den daraus resultierenden Beiträgen gibt, wenn die Prompts unterschiedliche Inhalte haben.

In einer zweiten Crowdsourcing-Studie werden Prompts evaluiert, die alle das Thema Lösungsvorschläge beinhalten, aber leicht unterschiedliche Ausrichtungen haben. Hier soll evaluiert werden, inwieweit die Spezifität der Prompts bei kollaborativer Reflexion relevant ist (Glogger u. a. 2009). Diese Studie bildet das gleiche Problem in einer simulierten Diskussionsumgebung ab, wie die vorherige Studie. Der Unterschied hinsichtlich des Aufbaus ist, dass in dieser Studie bereits ein Thema (das Problem) mit drei vordefinierte Antworten vorhanden sind. Da der Prompt für Lösungsvorschläge in der vorherigen Studie und in dieser Studie identisch ist, kann ermittelt werden, inwieweit das Vorhandensein von Antworten einen Einfluss auf die Wirkung von Prompts hat. Diese Studie verfolgt somit zwei Ziele: Zum einen eine Evaluierung der leicht unterschiedlicheren Formulierung, und zum anderen eine Analyse, inwieweit das Vorhandensein von Antworten einen Unterschied in der Wirkung eines Lösungs-Prompts hat.

Da die Auswirkung von Prompts auf geschriebene Texte evaluiert werden soll, muss der aus den Prompts resultierende Text inhaltlich analysiert werden. Hierfür eignet sich das Verfahren der qualitativen Inhaltsanalyse auf Basis eines Kodierschemas. In der vorliegenden Arbeit wird für diese Analyse das Kodierschema für kollaborative Reflexion von Prilla et al. verwendet (Prilla u. a. 2015). Auf diese Weise lässt sich ein Rückschluss auf die reflexiven Elemente eines Textes ziehen.

In den beiden Crowdsourcing-Studien wird nach Lösungsvorschlägen gefragt. Auf Basis der beiden Studien kann allerdings lediglich eine Aussage hinsichtlich der Häufigkeit von Lösungsvorschlägen in den Antworten gegeben werden. Eine Aussage zur Qualität der Lösungsvorschläge ist aber nicht möglich. Aus dem Grund wurde eine dritte

Crowdsourcing-Studie durchgeführt, in der Probanden die Lösungsvorschläge der beiden vorherigen Studien hinsichtlich unterschiedlicher aus der Literatur zur Reflexion abgeleiteter Kriterien bewerten. Das Ziel der Studie ist, dass eine neue Analysedimension für die beiden vorherigen Studien ermöglicht wird. Auf diese Weise können Aussagen getroffen werden, zu welchem Zeitpunkt (neues Thema vs. in einem Thema mit Antworten) oder bei welchem Prompt (Prompt für Lösungsvorschläge vs. Prompt für Erfahrungen) mögliche Unterschiede in der Qualität der Lösungsvorschläge auftauchen können.

Auf Basis der zwei durchgeführten Studien werden Empfehlungen für das Design von Prompts abgeleitet. Zusätzlich ergeben sich aus den Studien unterschiedliche Hinweise bzw. nicht eindeutige Ergebnisse, die zu neuen Hypothesen und weiteren künftigen Studien führen.

#### **1.4 Aufbau der Arbeit**

In diesem Kapitel wurde bereits die Motivation der vorliegenden Arbeit (Abschnitt 1.1) beschrieben. Ebenfalls findet sich in diesem Kapitel die zentrale Forschungsfrage der Arbeit (Abschnitt 1.2) sowie der Forschungsansatz (Abschnitt 1.3).

Kapitel 2 beschreibt die der Arbeit zugrunde liegende Literatur. Zunächst wird der Begriff der Reflexion (Abschnitt 2.4) bzw. kollaborativen Reflexion (Abschnitt 2.5) geklärt und im Kontext aktueller Literatur diskutiert. Zusätzlich wird das Konzept der Reflexion von anderen Konzepten wie bspw. Critical Thinking abgegrenzt, um den Begriff weiter zu konkretisieren (Abschnitt 2.7). Reflexion ist ein Konzept, welches informelles Lernen in der Praxis oder im beruflichen Kontext beschreibt, aber das in der täglichen Praxis unterschiedlichen Herausforderungen unterliegt (Abschnitt 2.6), welche die Reflexion erschweren (Abschnitt 2.8). Um diese Herausforderungen oder Erschwernisse zu überwinden, gibt es unterschiedliche Ansätze Reflexion zu unterstützen, die im selben Abschnitt erläutert werden. Um eine Unterstützung von Reflexion in Software abzubilden gibt es spezifische Modelle, wie bspw. das CSRL-Modell, was in Abschnitt 2.9 thematisiert wird. Diese Arbeit fokussiert sich auf Prompts als Unterstützungsform, welche in einem separaten Abschnitt ausführlich behandelt werden (Abschnitt 2.10). Ein möglicher praktischer Einsatzort von kollaborativer Reflexion sind sogenannte Communities of Practice (Abschnitt 2.1). Diese liegen ebenfalls im Fokus der vorliegenden Arbeit.

In Kapitel 3 wird das Konzept der vorliegenden Arbeit beschrieben. Dazu wird zunächst der Ansatz skizziert, wie Prompts als Unterstützungsform für Reflexion untersucht

werden sollen (Abschnitt 3.1). Anschließend erfolgt die Erstellung eines konkreten Konzepts, bei dem Ziele kollaborativer Reflexion auf Basis der Literatur identifiziert werden (Abschnitt 3.2). Zusätzlich werden spezifische Prompts zu den Zielen skizziert. Weiterhin wird in dem Kapitel erklärt, welche der Ziele in den einzelnen Studien dieser Arbeit untersucht werden sollen (Abschnitt 3.3). Da Prompts schriftliche kollaborative Reflexion beeinflussen sollen, müssen zur Überprüfung der Effektivität der Prompts die resultierenden Texte inhaltlich analysiert werden. Dazu wird das Konzept der Inhaltsanalyse erklärt und ein passendes Schema zur Auswertung von Reflexion ausgewählt und beschrieben (Abschnitt 3.4). Dies dient als Basis für die in dieser Arbeit durchgeführten Studien. Die Umgebung für die erste durchgeführte Studie ist eine Community of Practice Plattform mit selbst entwickelten Funktionen, um bspw. den Nutzern Prompts anzuzeigen (Abschnitt 3.5). Zusätzlich wird als Grundlage für die zweite Studie Crowdsourcing als Umgebung für Studien beschrieben (Abschnitt 3.6).

Kapitel 4 beinhaltet die erste Studie, bei der mehrere Ziele von kollaborativer Reflexion, die im Konzept beschrieben werden, über Prompts eingesetzt wurden. Die Umgebung der Studie war eine Community of Practice Plattform im Forschungsprojekt EmployID (Abschnitt 4.2), bei der die Prompts über eine eigens entwickelte Software verwaltet, angezeigt und dokumentiert wurden (Abschnitt 3.5). Die Studie wurde insgesamt in zwei Anwendungsfällen durchgeführt: bei Projektpartnern in Slowenien (Abschnitt 4.4) und Kroatien (Abschnitt 4.5). In beiden Anwendungsfällen erfolgt zunächst eine deskriptive Analyse der Daten, die eine Beschreibung der Nutzer der Community of Practice Plattform beinhaltet. Anschließend erfolgt eine quantitative Analyse, die zeigt welche Prompts wie oft gesehen bzw. beantwortet wurden. Ebenfalls wird für beide Anwendungsfälle eine qualitative Analyse der Daten durchgeführt. Die Ergebnisse beider Anwendungsfälle werden zunächst je Anwendungsfall und anschließend über beide Anwendungsfälle hinweg diskutiert (Abschnitt 4.6). Die Einschränkungen werden ebenfalls je Anwendungsfall und gemeinsam thematisiert (Abschnitt 4.7).

In Kapitel 5 erfolgt die zweite Studie, die sich aus drei unterschiedlichen Teilstudien zusammensetzt. In der ersten Teilstudie werden zwei konkrete Prompts für Lösungen bzw. Erfahrungen in einer Crowdsourcing-Studie evaluiert (Abschnitt 5.1). Die zweite Teilstudie in Abschnitt 5.2 untersucht hauptsächlich die Effekte von unterschiedlichen Formulierungen von Prompts, die alle versuchen Lösungsvorschläge anzuregen. Die beiden vorherigen Studien beinhalten Prompts zu Lösungsvorschlägen und analysieren hauptsächlich die Auswirkung auf die resultierenden Texte. Unklar ist, wie die Qualität der Lösungsvorschläge selbst beschaffen ist. Dies wird in der dritten Teilstudie analysiert, die in Abschnitt 5.3 beschrieben ist. Alle Studien werden in der beschriebenen Crowdsourcing-Umgebung durchgeführt (Abschnitt 3.5). Innerhalb der jeweiligen

Abschnitte für die Teilstudien wird die Methodik und die Auswertung der Daten beschrieben, sowie die Ergebnisse dargestellt und diskutiert.

In Kapitel 5.4 erfolgt eine übergreifende Diskussion der unterschiedlichen Studien dieser Arbeit, welche die Ergebnisse sowohl aus den Studien im realen Praxiseinsatz (siehe Kapitel 4) und aus dem Crowdsourcing-Bereich (siehe Abschnitte 5.1, 5.2 und 5.3) zusammenfasst und hinsichtlich des Konzepts diskutiert. Die Diskussion der Ergebnisse der unterschiedlichen Studien ist in mehrere Unterabschnitte untergliedert (Abschnitt 6.1), die einzelne Facetten der Ergebnisse thematisiert. Abschließend werden allgemeine Einschränkungen der vorliegenden Arbeit diskutiert, die noch nicht in den einzelnen Studien abgedeckt waren (Abschnitt 6.3).

Basierend auf den Ergebnissen dieser Studie werden in Kapitel 7 Empfehlungen für das Design von Prompts zur Unterstützung unterschiedlicher Aspekte kollaborativer Reflexion abgeleitet. Es werden zunächst spezifische Empfehlungen auf Basis konkreter Ergebnisse aus den Studien abgeleitet (Abschnitt 7.1). Anschließend werden Hypothesen aus den weniger eindeutigen Ergebnissen der vorliegenden Arbeit entwickelt und teilweise in einer ersten Skizze dargestellt (Abschnitt 7.2).

Abschließend wird die Arbeit in Kapitel 8 reflektiert und zusammengefasst.

## **2 Reflexion – Potenziale, Einsatzbereiche und Unterstützungsbedarfe**

In diesem Kapitel werden die Grundlagen für die vorliegende Arbeit thematisiert. Zunächst erfolgt eine Erläuterung von Communities of Practice als Rahmen der Arbeit und deren Bezug zu (kollaborativer) Reflexion. Anschließend wird der Begriff der Reflexion bzw. der kollaborativen Reflexion geklärt und die verschiedenen Einsatzbereiche von Reflexion thematisiert. Zusätzlich werden der Unterstützungsbedarf und unterschiedliche existierende Formen der Unterstützung, darunter Prompts, gezeigt.

### **2.1 Communities of Practice als Ort für Erfahrungsaustausch und Reflexion**

Communities of Practice sind ein Ort für kollaborative Reflexion, welcher oft innerhalb von Organisationen genutzt wird (Vince 2002), aber auch organisationsübergreifend bzw. –unabhängig existierend kann (Wenger, McDermott und Snyder 2002, 6).

#### **2.1.1 Was sind Communities of Practice**

Als Basis für Communities of Practice (CoP) wird an dieser Stelle das Konzept von Etienne Wenger angeführt (Wenger 1998; Wenger, McDermott und Snyder 2002). Wenger et al. bezeichnen als Community of Practice eine Gruppe von Personen, die eine gemeinsame Affinität zu einem Thema hat und sich regelmäßig austauscht, um das eigene Wissen und die Fähigkeiten zu diesem Thema weiterzuentwickeln (vgl. im Folgenden Wenger, McDermott und Snyder 2002, 1–22), oder neue Innovationen vorantreiben (Brown und Duguid 1991). Zusammen entwickelt diese Gruppe über Zeit eigenes Wissen, Richtlinien, Verfahrensweisen, Lösungsansätze. Diese Gruppe hat immer die Gemeinsamkeit über eine gemeinsame Praxis zu verfügen, was Wenger et al. bspw. mit Töpfern im alten Rom vergleichen, die sich zusammenschließen und nicht nur geschäftliche Interessen verfolgen, sondern auch zusammen lernen und Nachwuchs ausbilden (2002, 5). Oft werden Communities of Practice in Unternehmen genutzt, um den Austausch zwischen Gruppen, bspw. Fachkräfte einer Spezialisierung, zu ermöglichen (2002, 1–3 siehe einleitendes Beispiel), was darauf basiert, dass Personen gerne von Kollegen lernen (Hawkey 1995).

Es ist wichtig zu unterstreichen, dass eine Community of Practice keine Struktur ist, die ein Unternehmen erstellen kann, sondern eine Struktur ist, die sich über Zeit herausbildet (in der englischsprachigen Literatur wird von *emerge* gesprochen) (Thompson 2005; Wenger, McDermott und Snyder 2002). Eine Organisation kann allerdings der Community helfen zu wachsen, in dem sie ernst genommen wird und in das tägliche Geschäft einbezogen wird, bspw. in Entscheidungsfindungsprozesse (Wenger,

McDermott und Snyder 2002, 13). Ausbleibende Unterstützung impliziert nicht, dass die Community dann aufhört zu existieren (2002, 13). Eine zu starke oder zu schwache Involvierung der Organisation kann negative Effekte für die Community haben (Thompson 2005; Blunk und Prilla 2017a).

Sobald sich Personen mit ähnlicher Praxis zu einer Gruppe zusammentun und die Gruppe sich versucht über ihre Praxis (regelmäßig) auszutauschen beginnt eine Community zu entstehen. Diese Personen kennen oftmals die aktuell relevanten Themen bzw. wichtigen Herausforderungen und diese Personen werden daher oft zu Kernmitgliedern (im Orig. *core members*) der entstehenden Community (2002, 72). Es muss den Mitgliedern klar sein, dass eine Community of Practice mehr ist als nur ein persönliches Netzwerk, und dass es daher sinnvoll ist neue Personen hinzuzuziehen (2002, 72). Dies ist insbesondere der Fall, falls die neuen Personen weniger Wissen haben als die Kernmitglieder. Ein bekanntes Modell der Partizipation in Communities of Practice ist die *legitimate peripheral participation* von Lave und Wenger (Lave und Wenger 1991), welche besagt, dass neue Mitglieder zunächst in der Peripherie der Community of Practice ihre Teilnahme beginnen und je nach Aktivität in den Kern der Community of Practice vorrücken können. Dies nimmt je nach Art der Community und je nach Person durchschnittlich zwischen elf Monaten, bei durch die Organisation aktiv gestarteten Communities, und 17 Monaten bei natürlich entstehenden Communities in Anspruch (Borzillo, Aznar und Schmitt 2011). Nicht jedes Mitglied wird über die Laufzeit der Community bis in den Kern vorrücken, da dies von der Aktivität und den Aktivitäten selbst des jeweiligen Mitglieds abhängig ist (Lave und Wenger 1991). Der Kern einer Community umfasst ca. 10-15 % der Mitglieder während weitere 15-20 % der Mitglieder aktiv sind (Wenger, McDermott und Snyder 2002, 56). Der Rest der Mitglieder ist an der Peripherie angesiedelt und nur sehr selten aktiv (Wenger, McDermott und Snyder 2002, 56). Die Kernmitglieder der Community of Practice kümmern sich beispielsweise um übergeordnete Fragestellungen, wie beispielsweise der Organisation der Community, und nicht mehr nur das Lösen von fachspezifischen Problemen (Wenger, McDermott und Snyder 2002, 56). Die Teilnahme in der Peripherie einer Community of Practice beinhaltet beispielsweise das ausschließliche Lesen von Diskussionsbeiträgen oder das Zuhören in einer Diskussion, ohne einen eigenen Beitrag zu leisten (Wenger, McDermott und Snyder 2002, 56). Für Wenger et al. ist dies eine legitime Art der Teilnahme, weil auf diese Art neue Mitglieder in der Community von dem existierenden Wissen in der Community lernen können (Wenger, McDermott und Snyder 2002, 56). Gleichzeitig lernen sie etwas über die Community und den Umgang der Mitglieder miteinander.

Über die Teilnahme in der Community legitimieren die Personen in der Peripherie ihren Platz in der Community und werden so von den übrigen Mitgliedern akzeptiert und



können in den Kern vorrücken (Lave und Wenger 1991, 34–37). Die Peripherie und der Kern stellen somit unterschiedliche Ebenen an Aktivität in einer Community of Practice dar. Es gibt nicht zwingend nur einen Kern (Lave und Wenger 1991, 36). Die Mitglieder kennen sich gegenseitig und oft sind Communities of Practice an einen Ort gebunden, an dem man sich trifft (Johnson 2001). Persönliche Treffen sind eine passende Gelegenheit Erfahrungen oder Ideen auszutauschen (Tigelaar u. a. 2008).

Aus einer Übersicht (Blunk und Prilla 2017a) lassen sich die folgenden Gründe für die Teilnahme an einer Community of Practice bzw. die Barrieren für die Teilnahme an einer Community of Practice entnehmen. Vorteile an der Teilnahme ist die Möglichkeit voneinander zu lernen (Gray 2004) oder den eigenen Ruf in der Organisation zu fördern (Wasko und Faraj 2005), bspw. durch eigene Beiträge. Dazu kann man zusätzlich emotionale (Introne, Semaan und Goggins 2016), und fachliche Unterstützung von anderen erhalten, sowie voneinander lernen (Johnson 2001; Wenger, McDermott und Snyder 2002; Wenger 1998). Durch die Möglichkeit von anderen zu lernen, eignet sich eine Community of Practice für Anfänger, um an ein Thema herangeführt zu werden (Lesser und Storck 2001; Orlikowski 2008). Barrieren für die Teilnahme an Communities können vielfältig sein und damit anfangen, dass bspw. jemand lieber persönlich kommuniziert als elektronische Medien zu nutzen (siehe Abschnitt 2.1.3) (Ardichvili u. a. 2006). Ein geringes Selbstwirksamkeitsempfinden (im Orig. *self-efficacy*<sup>1</sup>) kann ein Grund für eine ausbleibende Teilnahme sein (Bandura 1993). Ein weiterer Grund kann das Bedürfnis sein, das eigene Wissen behalten zu wollen (Orlikowski 2008). Der Eindruck, die Praxis im Unternehmen nicht ändern zu können oder fehlender Anreiz dies zu tun, können weitere Gründe darstellen (Gambarotto und Cammozzo 2010; Gray 2004). In einem positiven Umfeld, wenn Wissen als öffentliches Gut geteilt wird, teilen die einzelnen Mitglieder ihr Wissen gerne (McLure Wasko und Faraj 2000).

Laut Blunk und Prilla (2017a; oder Titi Amayah 2013) gibt es aktuell wenig Forschung zu Communities of Practice in öffentlichen Verwaltungen (Callahan 2004), während der Fokus der Forschung eher auf privatwirtschaftlichen Unternehmen (Ardichvili, Page und Wentling 2003; Ardichvili u. a. 2006; Orlikowski 2008; Wenger, McDermott und Snyder 2002) oder im Bildungsbereich liegt (Gambarotto und Cammozzo 2010; Lisewski 2004; MacDonald 2008).

---

<sup>1</sup> Mit *self-efficacy* meint Bandura die eigene Überzeugung bzw. das eigene Empfinden eine Aufgabe erfolgreich durchzuführen, sodass ein bestimmtes Ergebnis produziert wird (Bandura 1977).

### 2.1.2 Informelles Lernen und Reflexion in Communities of Practice

Neben den erwähnten Vorteilen, wie Innovationen voran zu treiben (Brown und Duguid 1991), emotionale (Introne, Semaan und Goggins 2016) und fachliche Unterstützung von anderen zu erhalten (Johnson 2001; Wenger, McDermott und Snyder 2002; Wenger 1998), ist einer der wesentlichen Vorteile einer Community of Practice, dass man voneinander lernen kann (Johnson 2001; Wenger, McDermott und Snyder 2002; Wenger 1998; Lave und Wenger 1991; Eraut 2004).

Überwiegend handelt es sich beim Lernen in einer Community of Practice um informelles Lernen, sprich außerhalb des Kontextes eines formalen Trainings oder einer Fortbildung (Eraut 2004).

Reflexion ist eine Form informellen Lernens (siehe Abschnitte 2.1, 2.4 und 2.5) und dementsprechend oft eine Form, in der informelles Lernen in Communities of Practice auftritt (Vince 2002; Clouder 2000; Raelin 2002; Yang 2009; Stoszkowski und Collins 2017). Eine Übereinstimmung ist auch, dass, ähnlich zur Reflexion, bei Communities of Practice das Ziel ist, neue Ideen in die Praxis zu überführen (Vince 2002). Man bekommt gemeinschaftlich einen Sinn für die aktuellen Herausforderungen in der Praxis oder im Berufsfeld (Clouder 2000). Prilla merkt an, dass die Gruppengröße ein Faktor bei der Reflexion ist und in kleinen Gruppen die Situation aufkommen kann, dass niemand aktiv genug ist, um die Gruppe anzutreiben (vgl. im Folgenden Prilla 2014). In diesem Falle empfiehlt Prilla *Reflection Communities*, die, in Anlehnung an Communities of Practice, über Abteilungen oder Unternehmen hinweg als größere Gruppe zusammen reflektieren können. Prilla sieht weiterhin den Vorteil, dass diese Reflection Communities neue Mitglieder in den Unternehmen rekrutieren können, wenn diese realisieren, dass in der Community ein besseres Verständnis der Praxis erarbeitet wird. Dementsprechend erweitert er das Modell, siehe Abbildung 2-3, dahingehend, dass aus individueller Reflexion heraus in einer Reflection Community reflektiert werden kann, die dann ebenfalls Ergebnisse produziert, die zur Anwendung gebracht werden können.

In Studien wurde gezeigt, dass eine Blog-Umgebung als Ort für eine Community of Practice zur gemeinsamen Reflexion in der Ausbildung von Lehrern respektive von Studenten hilfreich ist (Yang 2009; Stoszkowski und Collins 2017). Dies wurde mit einem ähnlichen Medium in anderen Studien gefunden (Nicholson u. a. 2003). Eine zusätzliche Studie zeigt, dass Personen den fachlichen Austausch in Communities of Practice schätzen und diese nicht nur benutzen, um soziale Kontakte zu pflegen (McLure Wasko und Faraj 2000).

### 2.1.3 Benötigte IT-Unterstützung

In großen Unternehmen arbeiten oft mehrere Personen gleicher Spezialisierung an unterschiedlichen Orten, haben aber im Rahmen ihrer Tätigkeit keine natürlichen Gelegenheiten der Interaktion. Im Zuge der fortschreitenden Digitalisierung bilden diese Personen ebenfalls eine Community of Practice, wie bspw. im einleitenden Beispiel bei Wenger et al. dargestellt (Wenger, McDermott und Snyder 2002, 1–4). In dem Bezug wird oft von sogenannten *virtual communities* gesprochen, die sich über mehrere Orte oder Länder erstrecken (Johnson 2001). Die Communities existieren oft über das Internet, weil die Teilnehmer selten die Möglichkeit haben, sich persönlich zu treffen (Johnson 2001). Der Vorteil einer virtual community ist die Möglichkeit, dass die Personen sich dennoch leicht austauschen können, ohne sich persönlich zu kennen oder an einem Ort zu sein. Dementsprechend ist insbesondere für Communities of Practice, die in großen Organisationen oder im Internet existieren eine IT-Unterstützung notwendig.

## 2.2 Reflexion

In diesem Abschnitt wird eine kurze Übersicht über die kommenden Abschnitte zur Reflexion gegeben. Reflexion bezeichnet das Zurückkehren zu eigenen Erfahrungen und deren Neubewertung angesichts aktueller Kenntnisse mit der Zielsetzung neue Handlungsweisen für die Zukunft zu entwickeln (Boud, Keogh und Walker 1985). Die Basis ist immer das Lernen aus der eigenen Erfahrung heraus. Viele Berufe nutzen Reflexion als Bestandteil (siehe Abschnitt 2.3). Reflexion ist eine Form informellen Lernens, welches das Lernen außerhalb formaler Rahmenbedingungen, wie Trainings oder Schulungen bezeichnet (Eraut 2004).

Es wurde allerdings festgestellt, dass der Begriff der Reflexion oft lose genutzt wird (Walsh 2009, 389) bzw. dass viele Studien keine explizite Definition liefern, sich aber sehr oft für Reflexion auf die Arbeit von Schön beziehen (Baumer u. a. 2014). Traditionell hat sich Reflexion immer auf Einzelne bezogen (Dewey 1933; Schön 1983; Kolb 1984; Boud, Keogh und Walker 1985), was in der vorliegenden Arbeit als individuelle Reflexion bezeichnet wird. Die Arbeit zur individuellen Reflexion und den ersten Beschreibungen zur Reflexion in der Literatur wird in Abschnitt 2.4 behandelt. Wenn mehrere Personen gemeinsam reflektieren und Erfahrungen und Lösungen austauschen, dann wird dies in der vorliegenden Arbeit als kollaborative Reflexion bezeichnet. Eine Beschreibung des Begriffs und den dazugehörigen Grundlagen findet sich in Abschnitt 2.5.

## 2.3 Einsatzbereiche

Reflexion allein und mit Kollegen / Freunden / Familie wird in vielen unterschiedlichen Bereichen erfolgreich eingesetzt. Folgend ist eine Übersicht unterschiedlicher Anwendungen und Einsatzfelder.

Reflexion und kollaborative Reflexion ist ein Instrument für Fachkräfte in den unterschiedlichsten Berufen, um während der Tätigkeit über die eigene Praxis zu reflektieren oder um sich gegenseitig zu unterstützen. Reflexion wird in der Ausbildung von Gesundheits- und Krankenpflegern (Sandars 2009; Kim u. a. 2010; Platzer, Snelling und Blake 1997; Chirema 2007), Musikern (Johnston, Amitani und Edmonds 2005), Mediziner\*innen (Ryan 2010), Psychologen (Hettich 1990), Juristen (Anzalone 2010), Pharmazeuten (Wallman u. a. 2008), Sozialarbeiter\*innen (Murphy, Dempsey und Halton 2010), oder Lehrern (bspw. Craig 2010; Boerboom u. a. 2011; Davis 2006; Fleck und Fitzpatrick 2009; Nicholson u. a. 2003; Tigelaar u. a. 2008) verwendet. Hier fällt auf, dass dies überwiegend Berufe sind, bei denen Interaktion mit anderen eine große Komponente der täglichen Praxis ist und bei denen die Personen zu Fachkräften ausgebildet werden (Schön 1983).

Zwei jüngere Felder, in denen Reflexion erforscht wird, sind *Quantified Self* (Lupton 2016) oder *Personal Informatics* (Choe u. a. 2017), bei denen das Kernstück, das Sammeln von Daten über einen selbst ist, die man dann später analysieren kann. Zu diesen Daten gehören nicht nur Fitnessdaten, wie bspw. zurückgelegte Schritte bei einem Fitnesstracker, sondern Daten über den Schlaf, die eigene Stimmung, Finanzen oder den Energieverbrauch im Haushalt (Pirzadeh, He und Stolterman 2013). Da Reflexion lernen auf Basis eigener Erfahrungen ist, wird an dieser Stelle unterstrichen, dass man nicht über die angesammelten Daten an sich reflektiert, sondern über die mit den Daten verbundenen Erfahrungen (Fleck und Fitzpatrick 2010). Die Zielsetzung hinter der Aufzeichnung von Daten über einen selbst ist die persönliche Verbesserung bspw. bei Sportaktivitäten oder ein bewussterer Umgang mit Ressourcen, bspw. bei Finanzen oder Energieverbrauch. Es ist sowohl eine sinnvolle Visualisierung der Daten notwendig (Choe u. a. 2017) als auch eine Unterstützung bei der Interpretation der Daten (Puusaar, Clear und Wright 2017). In einer anderen Studie wurden Social-Media-Daten für den Nutzer zur Reflexion aufbereitet (Thomas u. a. 2018). Dazu wurden bspw. Bilder explizit für die Reflexion in Büchern, Bilderrahmen und Filmen neu aufbereitet. Die Autoren wollen damit über bloße Erinnerung hinausgehen und aktiv Reflexion zu den jeweiligen Erfahrungen ermuntern. Sie stellen fest, dass sich Social-Media-Daten sehr leicht zur Reflexion aufbereiten lassen

---

<sup>2</sup> Siehe <https://quantifiedself.com/> (Letzter Zugriff am 12.08.2019).

und die Teilnehmer die unterschiedlichen Formate gut nutzen können. Es wird aber hervorgehoben, dass Nutzer häufig bei unterschiedlichen Social-Media-Plattformen jeweils eine unterschiedliche Persönlichkeit darstellen, was durch die Art der Plattform bedingt ist und daher die Aufbereitung der Daten oft nur eine einseitige Reflexion ermöglicht.

Reflexion und das Sammeln von Daten über einen selbst wird unter anderem bei medizinischen Themen von Patienten eingesetzt. Ein Gebiet ist die Unterstützung des Selbstmanagements von Diabetespatienten (vgl. für diesen Abschnitt Mamykina u. a. 2016). In dieser Studie haben mehrere Probanden Zugriff auf ein Onlineportal gehabt, in dem sie einerseits Werte zum Blutzuckerspiegel eintragen und diese mit persönlichen Zielen für die eigene Verhaltensweise (bspw. mehr Wasser statt Soda trinken) verbinden konnten. Über Zeit konnten die Probanden so über das eigene Verhalten reflektieren und anhand der Daten nachvollziehen, welches Verhalten den Blutzuckerspiegel negativ beeinflusst. Mehrere Probanden haben daraufhin ihr Verhalten angepasst und dieses nach dem Ende der Studie beibehalten.

In einer Smartphone-App können Nutzer über eigene Erlebnisse anhand eines Fotos reflektieren (Isaacs u. a. 2013). Dieses Foto wird periodisch dem Nutzer angezeigt und dieser hat dann die Möglichkeit nach kurzer Reflexion einen Text zu dem Foto einzugeben und numerisch zu bewerten, wie glücklich man sich bei der Erinnerung fühlt. Die Studie von Isaacs et al. hat ergeben, dass die App das persönliche Wohlbefinden erhöhen kann. Die Autoren geben aber zu bedenken, dass es nachteilig ist, wenn die App Bilder mit negativen Assoziationen aufruft. Bei Episoden von Angstzuständen (im Orig. anxiety) bei Autisten kann technologie-unterstützte Reflexion in Form einer Smartphone-App helfen (Ferrario u. a. 2017). In der Studie hatten die Probanden eine App auf ihren eigenen Smartphones installiert, wodurch die Studie einen qualitativen Ansatz hat. Über extra Hardware konnten die Probanden sowohl positive als auch negative Momente in der App festhalten. Die App zeichnet beim Drücken des Knopfes auf der gesonderten Hardware den Ort und den Zeitpunkt auf. Über die App haben die Probanden dann einen Rückblick auf die letzten Wochen oder kürzere Zeiträume vornehmen können und reflektieren können.

In einer Anwendung zu körperlicher Fitness hat eine Smartphone-App versucht, Muster in dem Verhalten des Nutzers zu erkennen und den Nutzer darauf hin über Textnachrichten zum Reflektieren zu bewegen (Kocielnik u. a. 2018). Als Datenbasis dienten die Daten eines Fitnesstrackers. Falls Muster erkannt wurden, hat das System den Nutzer gefragt, wie er bspw. Tage mit niedriger Aktivität in der nächsten Woche vermeiden möchte. Das Ziel ist es, den Nutzer dazu zu bewegen sich eine Aktion fest

vorzunehmen. Das System kann anschließend überprüfen, ob ein quantifizierbares Ziel erreicht wurde. Einen Vorteil des Systems sehen die Autoren darin, dass ihr System den Nutzern hilft, neutral über ihre Aktivitäten zu reflektieren, ohne urteilend zu wirken und den Nutzer dazu zu drängen, bspw. mehr Sport zu machen. Die Dialoge, die den Nutzer in mehreren Schritten zum Nachdenken anregen, wurden von den Teilnehmern positiv bewertet.

Eine andere Studie hat gezeigt, dass Personen die gerne PC-Spiele spielen, dem Konzept der Reflexion gegenüber aufgeschlossen und der Meinung sind, dass Reflexion ihre Spielerfahrung vertieft (Mekler, Iacovides und Bopp 2018).

Zusammenfassend lässt sich feststellen, dass Reflexion einerseits eine sehr breite Anwendung in der Ausbildung zu unterschiedlichsten Berufen findet und andererseits im privaten Bereich an unterschiedlichen Stellen, wie bspw. Fitness oder Gesundheit, Hilfestellung leisten kann.

## **2.4 Individuelle Reflexion**

Dieser Abschnitt fokussiert sich auf individuelle Reflexion und beleuchtet die Arbeit von Dewey, Schön, Kolb und Boud et al., welche wichtige Grundlagen für individuelle Reflexion darstellen.

### **2.4.1 Reflexives Lernen nach Dewey**

Dewey ist einer der ersten Autoren, der Reflexion inhaltlich näher beschreibt und wird teilweise in späteren Werken häufig als Beginn für Reflexion genutzt (Hatton und Smith 1995; Boud, Cressey und Docherty 2006). Daher wird Dewey auch in dieser Arbeit als Grundlage genutzt.

Dewey bezeichnet reflexives Denken als „*Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends [...]*” (Dewey 1933, 9). Er bezeichnet Reflexion demnach als Auseinandersetzen mit Ansichten bzw. Annahmen, die jemand zu einem Thema hat. Er benennt Reflexion bzw. reflexives Denken als aus zwei Teilen bestehend: Erstens existiert eine Art Zweifel oder Ratlosigkeit (im Orig. *perplexity*), welche der Beginn des reflexiven Denkens darstellt und zweitens, die Bestrebung durch Suchen und Nachforschen diese Zweifel zu beseitigen (Dewey 1933, 12). Für den ersten Teil beschreibt er das Erleben einer Situation, die Fragen aufwirft und aufgrund derer man kurz inne hält (Dewey 1933, 12). Ohne diesen Auslöser geschieht keine Reflexion

(Dewey 1933, 15). Er betrachtet Reflexion als zielgerichtet und mit der Absicht die Situation, welche die Frage aufgeworfen hat, zu lösen (Dewey 1933, 14), was sich von üblichen Gedanken unterscheidet: Die Zielsetzung der Reflexion beschreibt er als: „*The nature of the problem fixes the end of thought, and the end controls the process of thinking*“ (Dewey 1933, 15).

Als notwendige Charaktereigenschaften, um reflexives Denken entwickeln zu können, nennt er Aufgeschlossenheit (im Orig. *openmindedness*, als frei von Vorurteilen (Dewey 1933, 30)), Interesse und Bereitschaft, Konsequenzen aus den Handlungen zu tragen (Dewey 1933, 33).

Er nennt Reflexion als insgesamt aus fünf verschiedenen Phasen bestehend, welche im Folgenden kurz beschrieben werden (vgl. im Folgenden Dewey 1933, 107–115):

- *Suggestion*: Hiermit sind Ideen gemeint, die auftauchen, sobald in einer Situation Zweifel oder Ratlosigkeit entsteht. Diese Ideen können dann untersucht werden. Falls mehrere Ideen gleichzeitig auftauchen, werden die Ideen gegeneinander evaluiert und eine davon wird ausgewählt.
- *Intellectualization*: Dies beschreibt den Wechsel von einem emotionalen Empfinden, wenn in einer Situation Zweifel oder Ratlosigkeit existieren, hin zu einem problembewussten Nachdenken zur Lösung dieser Situation. Dies bezeichnet ebenfalls die Identifikation des Problems und eines Lösungswegs.
- *Guiding idea*: Hiermit ist ein Einfall bzw. ein Gedanke gemeint, der plötzlich, ungeplant oder unfreiwillig kommt und eine erste Lösung für das Problem sein kann. Er bezeichnet dies als nicht inhärent intellektuell. Dies geschieht erst, sobald man sich mit dem Gedanken auseinandersetzt. Der Gedanke ist eine Art erste Hypothese, wie die Situation aufgelöst werden kann.
- *Reasoning*: Diese Phase bezeichnet das Weiterentwickeln des Gedankens aus der *Guiding Idea* Phase, in der bspw. weitere Einflüsse auf die Situation bzw. die Idee identifiziert werden.
- *Testing*: Diese Phase beinhaltet das Testen der Idee bzw. der Hypothese durch eine Handlung, ob diese funktioniert. Wenn das Problem gelöst wird und die übrigen Rahmenbedingungen zu dem Problem passen, wird dieses als gelöst betrachtet.

Wichtig ist, dass die Phasen keiner festen Reihenfolge unterliegen (vgl. für diesen Abschnitt Dewey 1933, 115–118), da bspw. eine Hypothese erdacht wurde, bevor das

Problem vollständig formuliert wurde. Ebenfalls führe jede Verbesserung der Idee zu neuen Beobachtungen. Dies kann somit eine neue Sequenz dieser Phasen hervorrufen. Auf diese Weise kann das Phasenmodell als zyklisch interpretiert werden. Als sechste Phase, die nicht im eigentlichen Modell enthalten ist, nennt er einen Blick in die Zukunft und in die Vergangenheit als Bestandteil von reflexivem Denken. Insbesondere bei dem Blick in die Vergangenheit sieht Dewey den Bezug zu den eigenen Erfahrungen.

Zusammengefasst lässt sich Reflexion nach Dewey als das Auseinandersetzen mit einer Situation verstehen, in der Zweifel oder Ratlosigkeit aufgekommen ist, die einen innehalten lässt und die Auseinandersetzung mit der Situation auslöst.

#### **2.4.2 Reflexion nach Schön**

Das Werk „*The Reflective Practitioner*“ von Donald Schön (Schön 1983) stellt für viele Wissenschaftler eine der zentralen Grundlagen für Reflexion bzw. reflexives Lernen dar (Baumer u. a. 2014). Als Basis dienen mehrere Konflikte, denen Fachkräften (im Orig. „*professionals*“) in den USA in den 1960er bis frühen 1980er Jahren gegenüberstehen (Vgl. für diesen Abschnitt Schön 1983, 9–17): Zum einen fällt es schwierig, den Standards ihrer Kunden gerecht zu werden und überfachliche Themen, wie gesellschaftliche Probleme, anzugehen. Zum anderen gab es Beispiele, in denen Fachkräfte, bspw. Ärzte und Anwälte, ihre berufliche Stellung und das Wissen für private Zwecke missbraucht haben. Zusätzlich beschreibt Schön die Problematik von Fachkräften, den Ursprung der eigenen Prozesse zu benennen und eigene Ansätze zur Problemlösung gegenüber anderen zu beschreiben oder anderen beizubringen. Schön schreibt den Berufen, die allgemeineres Wissen entwickeln einen höheren Status zu als Berufen, die spezifischeres Wissen entwickeln oder gar nur anwenden.

Die Herausforderung in der Praxis bei Fachkräften liegt darin, dass offen ist, welche Maßnahmen ergriffen werden müssen, falls sich eine Situation nicht direkt nach Handbuch lösen lässt, beispielsweise, weil sich die Situation nicht gleicht (Schön 1983, 28). Als Beispiel führt Schön dort Ärzte auf, deren Patienten Symptome aufweisen, die nicht eindeutig auf Krankheiten zurückzuführen sind, die dem behandelnden Arzt bekannt sind (Schön 1983, 28).

Schön beschreibt während einer Handlung implizit zu wissen, wie in einer Situation zu handeln ist, ohne dies konkret benennen zu können als *knowing-in-action* (Vgl. für diesen Abschnitt Schön 1983, 32–33). Schön hebt insbesondere die Dinge vor, die die Fachkraft in dem Moment nicht vollständig erklären könne. Dies sieht Schön als Anlass, dass Personen während der Tätigkeit innehalten und sich fragen, aufgrund welcher



Informationen man zu einer bestimmten Handlung gekommen ist. Hervorzuheben ist laut Schön, dass dies insbesondere während der Tätigkeit auftritt. Ein Beispiel für das Wissen während der Tätigkeit ist ein Fußballspieler, der beim Schuss während des Ausholens mit dem Fuß weiß, mit welcher Körperhaltung er den besten Erfolg in dieser Situation hat den Ball zum Tor zu bringen.

Dieses implizite Wissen (im Orig. „*tacit knowledge*“) beschreibt, dass man *mehr weiß als man erklären kann* (im Orig. „*knowing more than we can say*“, Schön 1983, 33). Schön beruft sich auf das Konzept von Polanyi, der als Beispiel das Erkennen eines bekannten Gesichts in einer Menschenmenge nennt, da man bei diesem Prozess nicht benennen kann, was zum Erkennen des Gesichts führt, aber man dennoch das Wissen hat, wer die betreffende Person ist (Polanyi 1966).

Dementsprechend benennt Schön *reflection-in-action* als Ansatz sich während einer Tätigkeit, in der man sich aktuell befindet, über die Tätigkeit und die eigene bisherige Ausführung dieser Tätigkeit zu hinterfragen (vgl. zu diesem Abschnitt Schön 1983, 34–38). Das Ziel dessen sei die Beibehaltung funktionierender Ansätze oder die Verbesserung bisher weniger erfolgreicher Ansätze die Tätigkeit durchzuführen. Als Beispiel führt Schön Jazz-Musiker auf, die während des Improvisierens Anpassungen des eigenen Spiels machen, um gemeinsam harmonischer zu klingen. Als Auslöser über die eigene Tätigkeit nachzudenken, wird die Überraschung (im Orig. „*surprise*“) genannt, die auftritt, wenn eine intuitiv durchgeführte Tätigkeit nicht zu dem erwarteten Ergebnis führt. Dies kann daran liegen, dass bspw. Situationen auftreten, die vorher unbekannt waren. Diese Überraschung über einen anderen Ausgang der durchgeführten Tätigkeit kann sowohl positiv, weil es besser als erwartet funktioniert hat, oder negativ, weil es weniger gut als erwartet funktioniert hat, sein. Der Fokus der Reflexion ist in dem Falle sowohl das Ergebnis der Handlung als auch die durchgeführte Handlung. Die Reflexion muss während einer Aktion durchgeführt werden, wobei Schön diese Aktionen als von der Praxis abhängig betrachtet und diese daher bereits nach Sekunden oder erst nach Wochen oder Monaten abgeschlossen sein können. Als Beispiel wird ein Anwalt angegeben, der nur wenige Sekunden reflektieren kann, bspw. vor Gericht, oder über Wochen und Monate reflektieren kann, wenn der Kontext ein längeres Gerichtsverfahren ist. Es wird über unterschiedliche Aspekte reflektiert, die einen unterschiedlichen Zeithorizont haben. Dementsprechend kann die Aktion länger andauern und damit die Reflexion ebenfalls (Schön 1983, 34–38). Hatton und Smith fügen später hinzu, dass *reflection-in-action* die schwerste Form der Reflexion ist (Hatton und Smith 1995).

Reflection-in-action kann auftreten, wenn man bei bekannten Ideen über die Tätigkeit reflektiert oder auch, wenn es für einen selbst eine neuartige Tätigkeit ist, bei der während

der Ausübung eine *theorie-in-action* entwickelt wird, die hilft die Tätigkeit adäquat durchzuführen (vgl. für diesen Abschnitt Schön 1983, 35–37). Dies taucht beispielsweise auf, wenn Kinder gleichmäßig und ungleichmäßig gewichtete Bausteine auf einer Stange balancieren sollen (Karmiloff-Smith und Inhelder 1974). Hier fangen die Kinder überwiegend bei gleichmäßig gewichteten Bausteinen mit dem geometrischen Mittelpunkt an und gehen dann bei ungleichmäßig gewichteten Bausteinen dazu über, jeden Baustein vorher auf den geometrischen Mittelpunkt zu prüfen. Es wird also während der Tätigkeit eine Theorie entwickelt, wie die Steine am besten zu platzieren sind und Kinder in der Studie reflektieren während der Ausführung der Tätigkeit über die Tätigkeit selbst. Schön benennt den Begriff des *reflection-in-practice*, welches sich im deutschen mit der ‚Praxis‘ eines Berufs übersetzen lässt. Konkret ist damit gemeint, dass Fachkräfte über ihre Praxis reflektieren, womit einzelne Fälle gemeint sind. Reflexion ist auch ein Korrektiv, um konkret über durch häufige Wiederholung eingeübte Abläufe nachzudenken, die bereits in *knowing-in-action* übergegangen sind (Schön 1983, 35–37).

Schön fasst zusammen, dass die Person bei der Anwendung von *reflection-in-action* selbst zum Forscher in dem eigenen Praxiskontext wird, da der aktuelle Kontext nicht mehr zum bisher erlernten oder bekannten passt und man somit überrascht wird (vgl. für diesen Abschnitt Schön 1983, 41–42). Schön sieht *reflection-in-action* als häufigen, aber außergewöhnlichen Prozess an. Der Umgang mit der Unsicherheit, was in einer neuen Situation konkret zu tun sei, werde von vielen aber auch als Anzeichen der Schwäche betrachtet, weil die Fachkraft in dem Moment keine Expertise vorweisen kann. Schön sieht *reflection-in-action* nicht als Form beruflichen Wissens akzeptiert.

In *The Reflective Practitioner* sind mehrere Beispiele genannt, in denen Personen, häufig in einer Mentor–Mentee-Konstellation, gemeinsam an einem Problem arbeiten (u. a. Architekturbeispiel (1983, 45 ff.) und Psychotherapiebeispiel (1983, 61 ff.)). Es werden neue Situationen als ähnlich, aber unterschiedlich zu bekannten Situationen eingeordnet, um das Einbeziehen bisheriger Erfahrungen zu ermöglichen (Schön 1983, 76–78). Die Fachkraft habe also ein *Repertoire* an Beispielen, Bildern, Verständnissen und Aktionen, auf die sie zurückgreifen kann. In anderen Beispielen hingegen wird angeführt, dass eine hohe kognitive Belastung in sozialen Interaktionen dazu führen kann, dass der betreffenden Person eine *reflection-in-action* nicht möglich ist (Schön 1983, 122). Zusätzlich sieht Schön die Reflexion bei Managern anders, da diese zusätzlich auf organisationales Wissen zurückgreifen können (Schön 1983, 129) bzw. die Reflexion allgemein in organisationalem Kontext stattfindet (Schön 1983, 140), und selten über die eigene Reflexion reflektiert wird.

Schön sieht weiterhin, dass reflection-in-action eine Art Experimentieren bzw. Testen von Hypothesen ist, mit dem Fachkräfte Ideen für neue Situationen entwickeln, testen und evaluieren (Schön 1983, 77–84). Werkzeuge für das effektive Reflektieren sind sowohl genutzte Artefakte (im Orig. „media“), Sprache und der eigene Erfahrungsschatz (Schön 1983, 144). Dennoch ist die Art zu reflektieren von Person zu Person und je nach Beruf unterschiedlich (Schön 1983, 145). Eine Kritik von außen an der Tätigkeit ersetzt aber keine eigene Selbstreflexion (Schön 1983, 152).

Schön stellt fest, dass das Nachdenken über eine Tätigkeit die Aktion selbst stoppen kann (vgl. zu diesem Abschnitt Schön 1983, 147–49). Beispielsweise gibt es Situationen, in denen innehalten und über die Tätigkeit reflektieren nachteilig oder gefährlich werden kann (z. B. im Straßenverkehr). Dementsprechend stellt Schön fest, dass es eine *reflection-on-action* gibt, welches erst nach der Tätigkeit durchgeführt wird. Betrachtet man den längeren Zeithorizont den Schön für manche Tätigkeiten vorgibt, bspw. längere Gerichtsverfahren, dann bieten sich in dem Zeitraum mehrere Möglichkeiten über die Tätigkeit nachzudenken, die noch aktuell läuft, ohne dass es in einer zeitkritischen Situation wie bspw. innerhalb einer Sitzung vor Gericht stattfinden muss. Hier beschreibt Schön, dass Fachkräfte lernen Pausen von wenigen Sekunden in einer Tätigkeit zur Reflexion zu nutzen, um die nächsten Schritte zu planen. Schön fasst zusammen, dass sich Denken und Handeln ergänzen, aber nicht gegenseitig blockieren. Dennoch bleibt offen, ob die Qualität der Reflexion bspw. für einen Anwalt, der zwischen zwei Gerichtsterminen mehrere Tage oder Wochen zur Reflexion Zeit hat, gleich den wenigen Sekunden an Zeit zur Reflexion ist, den bspw. ein Fußballspieler vor einem Elfmeter hat.

Schön sieht, insbesondere wenn eine Fachkraft mit einem Kunden, Patienten, oder Schüler interagiert, eine gemeinsame Komponente in der Reflexion, da diese mit der Fachkraft zusammenarbeiten müssen, um das jeweilige Problem zu lösen (Schön 1983, 156). Schön fokussiert sich in seiner Beschreibung von Reflexion immer auf allein reflektierende Fachkräfte, oder überwiegend auf die Kombination aus Fachkraft und Kunde, Patient, oder Schüler. Eine Beschreibung von Reflexion unter Fachkräften auf Augenhöhe findet nur eine oberflächliche Erwähnung (Schön 1983, 169). Die Ursache könnte darin liegen, dass Schön sich hauptsächlich mit reflection-in-action beschäftigt, und in seinen Beispielen Fachkräfte überwiegend allein einen Fall bearbeiten und nicht als Team von Fachkräften agieren. Dementsprechend gibt es während der Tätigkeit keine anderen Fachkräfte, mit denen man zusammen reflektieren könnte.

Auf organisationales Lernen bezogen, stellt Schön fest, dass Reflexion und das damit verbundene Hinterfragen von Aufgaben und Ansätzen eine bestimmte Belastung auf die Organisation ausübt (Schön 1983, 177). Die Organisation muss dafür bestimmte

Voraussetzungen mitbringen: Es muss Wert auf flexible Prozeduren oder qualitative Wertschätzung von komplexen Prozessen gelegt werden (im Orig. „[...] a high priority on flexible procedures, [...] qualitative appreciation of complex processes [...]“ (Schön 1983, 176)). Zusätzlich muss die Organisation Raum schaffen, um Konflikte zwischen Werten und Zwecken zu adressieren (im Orig. „[...] must make a place for attention to conflicting values and purposes.“ (Schön 1983, 177)).

Zusammengefasst beschreibt Schön reflection-in-action als ein Prozess, mit dem man seine Tätigkeit überdenkt, nachdem der Ausgang einer Aktion in der Tätigkeit für einen überraschend war. Schön beschreibt insgesamt mehrere Elemente, die zum Prozess der reflection-in-action gehören: Es taucht eine Überraschung auf, dass eine Aktion anders verläuft, als man es erwartet hat; man erstellt eine Hypothese, wie man diese Aktion in Zukunft angehen kann; man führt die neue Version der Aktion durch. Dieser Prozess hilft Fachkräften, aus eigenen Tätigkeiten für die Zukunft zu lernen. Schön unterscheidet reflection-in-action von reflection-on-action, welche lediglich nach einer Tätigkeit ausgeführt wird und nicht, während die Fachkraft noch in der Ausführung der Tätigkeit, die sich über längere Zeiträume erstrecken kann, befindet.

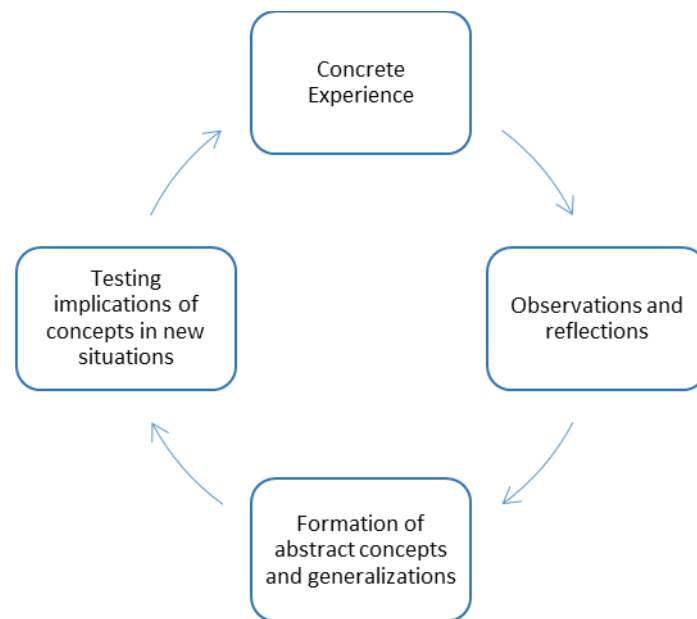
### **2.4.3 Experiential Learning nach Kolb**

Während Schön Reflexion als Lernen auf Basis aktueller Erfahrungen überwiegend in beruflichen Situationen beschreibt, widmet sich Kolb hauptsächlich dem Thema Lernen aus Erfahrung und bezieht seine Arbeit häufig auf den schulischen bzw. universitären Bildungsbereich.

Kolb sieht Lernen als zentralen Aspekt im Leben, da seiner Aussage nach die Welt schneller wird, die Menschen dadurch schneller auf neue Einflüsse reagieren müssen und Organisationen verstärkt in Konkurrenz zueinanderstehen (Kolb 1984, 1–4). Eine der frühen Herausforderungen in der Lehre war, dass sich die Forschung darauf konzentriert hat *was* gelehrt werden soll und weniger darauf *wie* es gelehrt werden soll (Kolb 1984, 7). Daraufhin skizziert Kolb die Vorarbeiten anderer, um die Entwicklungen in der Pädagogik aufzuzeigen, und stützt sich für die Experiential Learning Theorie insbesondere auf die Vorarbeiten von Dewey, Lewin und Piaget (Kolb 1984, 1–19, siehe insbesondere Grafik 1.2, 17).

Von Lewin entleiht Kolb sich das Konzept der *T-Groups* (*training groups*) (Lewin 1951; zitiert in Kolb 1984), in denen zu Lerneinheiten diskutiert wurde, und bezeichnet diese Gruppen als eine der wichtigsten Entdeckungen im Lehr- Lernbereich im 20. Jahrhundert (Kolb 1984, 8–10). Kolb beschreibt einen Lernzyklus für erfahrungsbasiertes Lernen,

unter anderem nach den Vorarbeiten von Lewin, mit vier Schritten (siehe Abbildung 2-1; D. A. Kolb 1984, 21–22; D. A. Kolb and Fry 1975):



**Abbildung 2-1: Experiential learning Zyklus nach Kolb und Fry (Kolb und Fry 1975, Figure 1)**

Der Lernzyklus (siehe Abbildung 2-1) beginnt mit der Erfahrung selbst, welche die Basis für Beobachtung und Reflexion ist. Im dritten Schritt wird aus der Beobachtung und Reflexion eine neue Theorie entwickelt, welche in neuen Situationen ausprobiert wird (Schritt vier), was dann zu neuen Erfahrungen führt und somit den Kreis schließt. Zu beachten ist, dass in diesem Modell bereits von Reflexion gesprochen wird, wobei die Reflexion hier nur die reine Interpretation der Erfahrung impliziert. Kolb hebt hervor, dass eine Besonderheit des Modells die Verankerung an der Erfahrung im Hier und Jetzt ist (im Orig. „[...] here-and-now concrete experience [...]“), was ähnlich wie bei Schön auf eine Theorieneuevaluierung während einer Tätigkeit schließen lässt (Kolb 1984, 22). Kolb meint damit, dass das aktuelle Erleben einer Erfahrung zum Lernen führt und nicht der Rückblick auf eine Erfahrung (Kolb und Kolb 2018). Kolb und Kolb berufen sich in dem Artikel auf Dewey<sup>3</sup>, nach dem ein Lerner wirklich in einer Situation feststecken muss, damit dies zu einer Reflexion führt. Kemmis führt im gleichen Werk wie Boud et al. aus (Boud, Keogh und Walker 1985), dass es immer einen Grund in Form einer Art Problem gibt, der einen innehalten lässt und förmlich zur Reflexion treibt (Kemmis 1985, 141).

---

<sup>3</sup> In dem Artikel wird Dewey nur genannt, aber nicht korrekt zitiert!

Kolb beschreibt das Lernmodell von Dewey hinsichtlich Lernen aus Erfahrung stark ähnlich der Arbeit von Lewin, aber hebt hervor, dass der Prozess des Lernens aus Erfahrung deutlicher gemacht wird (Kolb 1984, 23).

Kolb fasst die Arbeit von Piaget<sup>4</sup> so zusammen, dass die Intelligenz durch die eigene Erfahrung beeinflusst wird, und die wiederum durch die Interaktion mit der Umwelt definiert wird (Kolb 1984, 11–12). Diese Interaktion mit der Umwelt als Ansatz zu Lernen bezeichnet Kolb als ähnlich zu den Ansätzen von Lewin und Dewey (Vgl. für diesen Abschnitt Kolb 1984, 23): Den Ansatz beschreibt Kolb als Wechselwirkung zwischen dem Prozess der Übernahme (im Orig. „accommodation“) von Konzepten und dem Prozess Erfahrungen in eigene Konzepte einzusortieren (im Orig. „assimilation“). Wenn der erste Prozess überwiegt, dann ahmt man seine Umgebung nach, und wenn der andere Prozess überwiegt, so lernt man im Spiel, der Betrachtung eigener Konzepte ohne Rücksicht auf die Umgebung bzw. Realität. Lernen nach Piaget findet im Wechselspiel der beiden Prozesse statt und durchläuft auf dem Weg vom Kind zum Erwachsenen unterschiedliche Stadien (Piaget 1976).

Kolb sieht *experiential learning* als Baustein für selbstgesteuertes (im Orig. „*self-directed*“) Lernen (Kolb 1984, 18). Dies kann für Kolb ein Baustein für *Karriereentwicklung* (im Orig. „*career development*“) und *lebenslanges Lernen* (im Orig. „*lifelong learning*“) sein. Kolb unterstreicht das *experiential learning* keine Sammlung von Techniken zur Anwendung, sondern ein Programm zur Ausrichtung des persönlichen Lebens bzw. von sozialen Systemen ist (im Orig. „[...] is not a series of techniques to be applied in current practice but a program for profoundly recreating our personal lives and social systems.“) (Kolb 1984, 18), das man Anwenden kann, sondern eher als Programm zur Gestaltung von Lehre und Leben ist.

Kolb fasst die Arbeit von Lewin, Dewey und Piaget zusammen, wonach Lernen ein Prozess ist, bei dem Konzepte sowohl durch Erfahrung gebildet werden als auch kontinuierlich durch Erfahrung weiterentwickelt werden (Kolb 1984, 26). Kolb sieht eine Idee nicht als abgeschlossene Entität, sondern immer als Produkt aus Gedanken und Erfahrungen. Dementsprechend sieht Kolb keine zwei Gedanken als identisch an. Jeder lernt kontinuierlich aus den eigenen Erfahrungen (Kolb 1984, 27). Dies geschieht jedoch nicht auf der grünen Wiese, da jeder Lernende eigene Vorerfahrungen hat (Kolb 1984, 28). In späteren Artikeln wird das als *re-learning* bezeichnet, bei dem man bisher gelerntes Wissen bei neuen Erfahrungen erweitert bzw. anpasst (Kolb und Kolb 2018). Auf dieser Basis und den Vorarbeiten von Lewin, Dewey und Piaget erweitert Kolb den

---

<sup>4</sup> Im Text von Kolb 1984 wird keine spezifische Quelle für Piaget im Text genannt.

obigen Lernzyklus (siehe Abbildung 2-1). Nach diesem benötigt ein Lernender vier Fähigkeiten: „*Concrete Experience abilities (CE)*“, „*Reflective Observation abilities (RO)*“, „*Abstract Conceptualization abilities (AC)*“, und „*Active Experimentation abilities (AE)*“ (Kolb 1984, 30). Kolb sieht diese Fähigkeiten im Gegensatz zueinander und fragt, wie jemand gleichzeitig handeln und reflektieren kann (im Orig. „*How can one act and reflect at the same time?*“) (Kolb 1984, 30). Kolb sieht die Fähigkeiten als Gegensätze an, bei denen sich der Lernende entscheiden muss, welche Fähigkeit man in welcher Situation einsetzt. Diese vier Fähigkeiten setzt Kolb jeweils als Endpunkte auf zwei Dimensionen, von denen die erste Dimensionen über die Endpunkte „*concrete experience*“ und „*abstract conceptualization*“, und die andere über die Endpunkte „*active experimentation*“ und „*reflective experimentation*“ verfügt (Kolb 1984, 30–31). Beim Lernen sieht Kolb nicht alle Dimensionen als gleichmäßig ausgeprägt und einzelne Dimensionen können das Lernen dominieren. Bei Schön ist die Definition von *reflection-in-action* dahingehend schwierig, dass einzelne Zeiteinheiten nicht deutlich sind (Schön 1983, siehe Abschnitt 2.4.2). Es kann argumentiert werden, dass Denken und Handeln so abgestuft betrachtet wird, dass zwar innerhalb einer Aktion reflektiert wird, aber Denken und Handeln separat sind, was die Aussage von Kolb bestätigt.

Zusammenfassend definiert Kolb *experiential learning*<sup>5</sup> an zwei Stellen wie folgt:

„[experiential] *Learning is the process whereby knowledge is created through the transformation of experience*“ (Kolb 1984, 38)

„[experiential] *Learning, the creation of knowledge and meaning, occurs through the active extension and grounding of ideas and experiences in the external world an through internal reflection about the attributes of these experiences and ideas*“ (Kolb 1984, 52)

Damit unterstreicht Kolb, dass Wissen etwas ist, was man aus den Erfahrungen erlangt und nicht abgeschlossen ist, sondern kontinuierlich neu erstellt bzw. angepasst wird, und das Wissen keine Entität ist, die man so aufnimmt (Kolb 1984, 38). Zudem sieht Kolb Lernen als Vorgang, der Erfahrungen transformiert.

Mit dieser Definition liegt Kolb sehr dicht an der Arbeit von Schön und das Lernen auf Basis von Erfahrung, was Kolb skizziert hat, lässt sich in der Argumentation von Schön wiederfinden (siehe Abschnitt 2.4.2). Dieser sieht Erfahrung als Ausgangspunkt, um bei

---

<sup>5</sup> Im Buch verweist Kolb in Fußnote 3 (Kolb 1984, 38) darauf, ab dem Zeitpunkt nur noch von *Learning* im Sinne von *experiential learning* zu schreiben. Zum besseren Verständnis wird an diesen Stellen in den beiden Definitionen das Wort „experiential“ eingefügt.

einer Überraschung die eigene Theorie über den Sachverhalt anzupassen, neu auszuprobieren und so das eigene Verhalten in die eine oder andere Richtung anzupassen. Dementsprechend lassen sich Kolb und Schön differenzieren, da Schön von überraschenden Situationen spricht, in denen man beginnt zu reflektieren, und Kolb ganz allgemein von Lernen aus Erfahrung spricht, ohne dass es einen bestimmten Anlass gibt.

Kolb stellt fest, dass jeder Lerner unterschiedlich ist, und gibt zu bedenken, dass man diesbezüglich nicht in Stereotypen verfallen darf (Kolb 1984, 62–63). Jeder Mensch hat seiner Ansicht nach unterschiedliche Charaktereigenschaften, die allerdings nicht fix sind, sondern eher stabilen Zuständen gleichen, sich also im Verlauf des Lebens ändern können (Kolb 1984, 63). Ein Beispiel für die unterschiedlichen Lernansätze sind bspw. analytisches Denken bei Ingenieurstudenten und kreatives Denken bei Kunststudenten (Kolb 1984, 166). Dementsprechend verfügt jeder über unterschiedliche Lernansätze, die ebenfalls nicht fix sind und sich ändern können bzw. je nach Lernzweck unterschiedlich aussehen können (Kolb 1984, 95–98). Man neigt aber dazu, die Lernansätze beizubehalten bzw. zu wiederholen, mit denen man in der Vergangenheit Erfolg hatte (Kolb 1984, 95–98). Zu diesen Lerntypen hat Kolb einen *Learning Style Inventory*-Fragebogen entwickelt und erforscht wie Lerner konkret aus Erfahrungen lernen, um zu analysieren, welche Schwerpunkte Personen in bestimmten Disziplinen setzen (Kolb 1984, 161 ff.). Dieser Fragebogen wurde kontinuierlich weiter beforscht (bspw. Kolb und Kolb 2005) und weiterentwickelt (Kolb 2007). Kolb beschreibt die unterschiedlichen Lerntypen als unterschiedliche Wege durch den Lern-Zyklus (Kolb und Kolb 2018).

Kolb spricht bei experiential learning von unterschiedlichen Ansätzen, etwas zu lernen, was zuvor bereits bzgl. den Dimensionen beschrieben wurde. Kolb beschreibt den Lernansatz „*apprehension transformed by intention*“, mit einem Beispiel eines Billard-Spielers der nach seinem Spielzug über selbigen nachdenkt und nennt dies reflexives Lernen (Kolb 1984, 65). Der Spieler in Kolbs Beispiel denkt im Spiel einerseits über seinen eigenen Spielzug nach, aber hat die Möglichkeit bei anderen zuzuschauen, um von deren Spielzügen zu lernen. Es fehlt das bei Schön erwähnte Überraschungselement (siehe Abschnitt 2.4.2; Schön 1983). Zusätzlich wird bei Kolb nur davon gesprochen, dass der Lernende sich Inspiration bei anderen holt (Kolb 1984, 65). Eine explizite Kommunikation zwischen zwei Personen zu Lernzwecken wird nicht beschrieben. Später präzisiert Kolb reflexives Lernen als „*reflective observation*“, mit der Kolb das Betrachten unterschiedlicher Sichtweisen und das Interpretieren von Situationen meint (Kolb 1984, 68–69). Kolb schreibt dieser einen Fokus auf die eigenen Gedanken und Gefühle zu (Kolb 1984, 68–69). Dies ist für Kolb einer der Lerntypen, was impliziert, dass es Personen gibt, die verstärkt reflexiv lernen und eben solche, die eher anders



lernen. Eine detaillierte Beschreibung, wie Reflexion abläuft, gibt Kolb nicht an, was Boud et al. anmerken (Boud, Keogh und Walker 1985, 13).

Kolb sieht Lernen als lebenslangen Prozess, der sowohl schulisch bzw. universitär als später im Beruf stattfindet (Kolb 1984, 128; Kolb 2007; Kolb und Kolb 2018).

In dem Modell zu *experiential learning* beleuchtet Kolb zudem detailliert kognitive Prozesse, die beim Lernen im Gehirn ablaufen, und betrachtet unterschiedliche Entwicklungsstufen im Leben, bspw. sowohl das Erwachsenwerden als auch der weitere Verlauf im Leben und der spätere berufliche Werdegang. Diese Bereiche sind nicht wiedergegeben, da die kognitiven Prozesse auf der Ebene für das weitere Verständnis von Reflexion nicht notwendig sind.

Zusammenfassend lässt sich feststellen, dass bei Kolb, ähnlich wie bei Schön (siehe Abschnitt 2.4.2), die Erfahrung die Ausgangsbasis für Lernprozesse darstellt. Anders als bei Schön ist eine Überraschung kein wichtiges Element, welches Denkprozesse anstößt.

#### **2.4.4 Reflexives Lernen nach Boud et al.**

Boud et al. beginnen mit der Prämisse, dass reflexives Lernen bzw. der Prozess der Reflexion in den bisherigen Werken von Dewey und Kolb nicht detailliert genug beschrieben wurde (Boud, Keogh und Walker 1985, 13) und in der schulischen Ausbildung oft vernachlässigt wird (Boud, Keogh und Walker 1985, 7–9). Typische Lerninhalte, die didaktisch *erfahrungsbasiert* (im Orig. „experience-based“) durchgeführt werden, sind Praxiswochen, bei denen die Studenten für einen bestimmten Zeitraum einer praktischen Tätigkeit nachgehen (Boud, Keogh und Walker 1985, 8–11). Boud et al. beschreiben insgesamt drei Phasen, die die Studenten für die Praxiswoche durchlaufen: Vorbereitung, Durchführung und Nachbereitung, wobei Boud et al. anmerken, dass in solchen Phasen für die Studenten oft eine mentale Trennung zwischen theoretischen Inhalten aus dem Unterricht und dem in der Praxis Beobachteten oder Durchgeführten liegt. Die Nachbereitung erfolge dann aufgrund der angefertigten Notizen, welche die Studenten häufig anlegen, um die Flut der neuen Informationen zu bewältigen.

Boud et al. halten fest, dass Lerner nur über die eigenen Erfahrungen reflektieren können (Boud, Keogh und Walker 1985, 11). Dementsprechend hätten Lehrer nur Zugriff auf die Informationen, die bei der Reflexion ausgesprochen oder visualisiert werden. Boud et al. sehen Reflexion als zielgerichteten Prozess an, der über bloßes Nachdenken hinausgeht. Zusätzlich sehen Boud et al. Reflexion und Gefühle an sich stark verwoben, wobei

positive Gefühle den Lernprozess unterstützten können und negative Gefühle sowohl die Wahrnehmung verzerren als auch das Lernen behindern oder verhindern können.

Als Grundlage für den Lernprozess berufen Boud et al. sich auf den vierstufigen Lernprozess für experiential learning von Kolb (siehe Abschnitt 2.4.3, Kolb 1984; Kolb and Fry 1975). Zusätzlich stellen Boud et al. fest, dass dieser den Prozess der Reflexion nur oberflächlich beschreibt (Boud, Keogh und Walker 1985, 13).

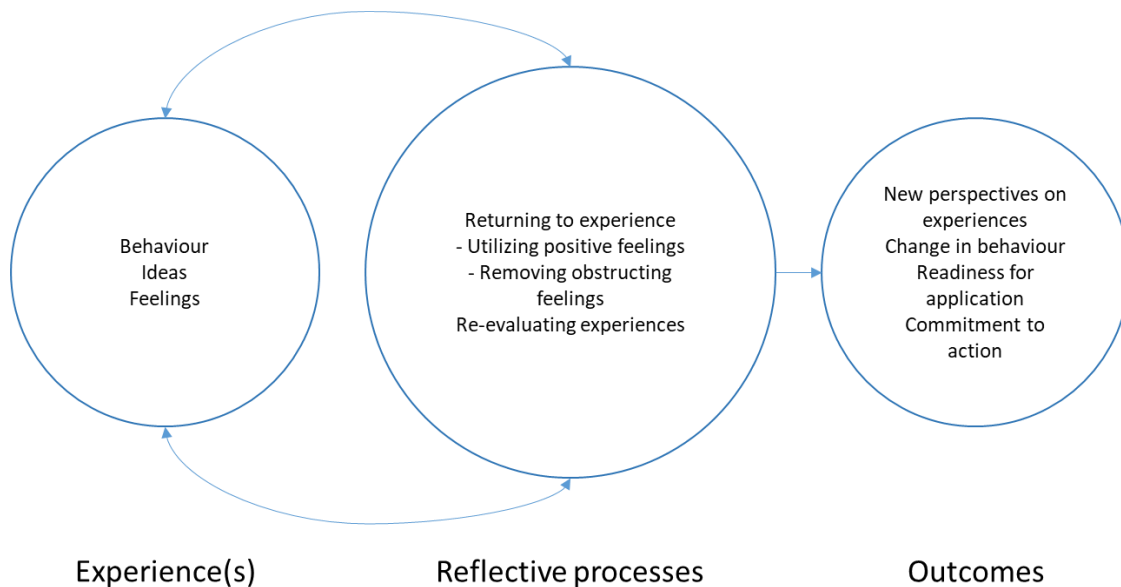
Für das eigene Modell zu reflexivem Lernen beschränken sich Boud et al. auf Reflexion als aktives, absichtliches und zielgerichtetes Lernen (Boud, Keogh und Walker 1985, 18–19). Boud et al. beschreiben die Erfahrung selbst und die dazugehörige reflexive Handlung als die beiden Hauptkomponenten von Reflexion. Eine Erfahrung beschreiben Boud et al. als vollständige Reaktion auf eine Situation einschließlich der eigenen Gedanken, Gefühle, Handlungen, Reaktionen und/oder Schlussfolgerungen. Die Situation könne geplant oder ungeplant, privat, beruflich oder in der Ausbildung sein. Die Erfahrung charakterisieren Boud et al. als recht komplex und aus mehreren kleineren spezifischeren Erfahrungsbestandteilen bestehend. Die zweite Komponente, die Reflexion, werde als Nachdenken und Neuevaluieren der Situation mit Zielsetzung neuen Verständnisses dargestellt. Das Lernen stamme so aus der Verarbeitung der Erfahrung. Boud et al. legen Wert darauf, dass Reflexion bewusst abläuft, da der Lernende bei unterbewusster Reflexion keinen Einfluss auf Entscheidungen über den Lernprozess hat. Lernende reflektieren allerdings nicht über jede Erfahrung (Boud, Keogh und Walker 1985, 20).

Als Ursache für Reflexion wird immer ein Stimulus beschrieben (Boud, Keogh und Walker 1985, 19), wobei Boud et al. sich dazu auf den von Boyd und Fales beschriebenen „*discomfort*“ beziehen, mit dem die Unzufriedenheit mit einem selbst in einer gegebenen Situation gemeint ist (Boyd und Fales 1983), oder die in Dewey beschriebene Verwunderung (im Orig. „*perplexity*“, Dewey 1933). Boud et al. beschreiben, dass Reflexion aus positiven Gefühlen entstehen kann, bspw. falls man erfolgreich eine Aufgabe erledigt hat (Boud, Keogh und Walker 1985, 19).

Im Gegensatz zu Dewey betonen Boud et al. den Faktor der gefühlsbezogenen (im Orig. „*affective*“) Aspekte der Reflexion und deren begünstigender oder einschränkender Einfluss auf Reflexion (Boud, Keogh und Walker 1985, 21).

Boud et al. unterscheiden zwischen *tiefem* (im Orig. „*deep*“) und *oberflächlichem* (im Orig. „*surface*“) Lernen (Boud, Keogh und Walker 1985, 24; Entwistle und Ramsden 1983). Ersteres umfasse die Zielsetzung das Gelernte wirklich verstehen zu wollen und verknüpft unterschiedliche Wissensgebiete, Letzteres sei umgangssprachlich „auswendig

lernen“ von Fakten ohne Zielsetzung des Verständnisses. Sie verorten Reflexion bei dem tiefen Lernen (Boud, Keogh und Walker 1985, 24).



**Abbildung 2-2: Reflexives Lernen nach Boud et al. (1985, 36, Figure 3)**

Für ihr Modell der Reflexion, siehe Abbildung 2-2 (links), sehen Boud et al. die Erfahrung des Lerners als Ausgangspunkt. Dieser Ausgangspunkt ist die Gesamtheit der eigenen Erfahrungen, des Verhaltens in den Situationen, der eigenen Ideen und Gefühle (Boud, Keogh und Walker 1985, 20). Diese Gesamtheit der eigenen Erfahrungen bestimmt, wie man selbst die Welt sieht und ist bei jeder Person unterschiedlich (Boud, Keogh und Walker 1985, 21). Die eigentliche reflexive Phase, siehe Abbildung 2-2 (Mitte), läuft in drei Etappen ab (Boud, Keogh und Walker 1985, 26–36):

In der ersten Etappe greift man auf die eigene jeweilige Erfahrung zurück, über die man reflektieren möchte (vgl. für diesen Abschnitt Boud, Keogh und Walker 1985, 26–36). Dazu versuche man sich auf das zu konzentrieren, was während der Erfahrung wirklich passiert ist und wie man sich gefühlt hat. Es könne helfen die Erfahrung aufzuschreiben (siehe Abschnitt 2.8). Während dieser Phase solle man darauf achten noch nicht über die Erfahrung zu urteilen, sondern eher möglichst detailliert die Erfahrung neu zu durchleben. Man könne aber die entsprechenden Interpretationen und Beurteilungen, die man in der jeweiligen Situation getätigt hat, erfassen. Zusätzlich solle man in diesem Schritt seine Gefühle erfassen, die man während der Erfahrung gemacht hat.

In der zweiten Etappe ist es wichtig, die mit der Erfahrung assoziierten Gefühle zu analysieren (vgl. für diesen Abschnitt Boud, Keogh und Walker 1985, 26–36): Positive

Gefühle könne man nutzen, um motiviert in die Reflexion einzusteigen. Negative Gefühle könnten eine Barriere für Reflexion darstellen und die Sicht auf die Erfahrung verzerren oder gar ein falsches Bild von der Erfahrung erzeugen. Daher solle man diese nach Ansicht von Boud et al. entfernen, um eine möglichst rationale Sicht auf die Erfahrung zu bekommen und um mit dem Lernprozess weiterzumachen. Boud et al. nennen dafür unterschiedliche Ansätze, wie das Niederschreiben von Gefühlen oder Gesprächen mit Dritten.

Die dritte Etappe ist die Neu-Evaluierung der eigenen Erfahrung vor dem Hintergrund der Zielsetzung der Reflexion (vgl. für diesen Abschnitt Boud, Keogh und Walker 1985, 26–36). Wichtig ist Boud et al., dass man nicht direkt von der Erfahrung in die Neu-Evaluierung springt, sondern erst die Erfahrung Revue passieren lässt und sich der assoziierten Gefühle vergegenwärtigt. Boud et al. nennen vier Aspekte, die die Reflexion unterstützen: *Assoziation* (neue Daten zu der bekannten Situation hinzufügen), *Integration* (Verbindungen zwischen den Daten finden), *Validation* (Überprüfen der Ideen) und *Aneignung* (im Orig. „appropriation“) des neuen Wissens. Diese vier Aspekte sollen keine Schritte darstellen, die man durchlaufen muss, sondern Elemente der Reflexion an sich, die teilweise aufeinander folgend sein können. Diese vier Aspekte werden im Folgenden beschrieben:

Bei der *Assoziation* werden die Gedanken und Gefühle, die während der Erfahrung aufgetreten sind, mit den Ideen verbunden, die während der Reflexion aufgetreten sind (vgl. für diesen Abschnitt Boud, Keogh und Walker 1985, 26–36). Er fasst zusammen, dass das Herstellen dieser Verbindungen der zentrale Prozess beim Lernen sei. Es könne einem auffallen, dass alte Annahmen zu einem Sachverhalt nicht länger valide sind bzw. nicht länger zu den neuen Ansichten passen. Eine Technik diesen Aspekt direkt zu fördern ist das Brainstorming (Osborn 1953).

Während der *Integration* werden die Assoziationen geprüft, inwieweit sie nützlich sind (Boud, Keogh und Walker 1985, 26–36). Die Integration hat zwei Elemente: Zum einen werden die Charakteristiken der neuen Verbindungen (im Orig. „the nature of relationships“; Boud, Keogh und Walker 1985, 32) analysiert und zum anderen werden Schlussfolgerungen aus dieser Analyse gezogen.

In der *Validation* werden die neuen Ideen und Gedanken geprüft, ob sie konsistent sind, ob sie zu existierendem Wissen passen (vgl. im weiteren Abschnitt Boud, Keogh und Walker 1985, 26–36). Falls festgestellt wird, dass die Idee inkonsistent zum bisherigen Wissen ist, kann diese dann neu evaluiert werden. Man kann neue Ideen beibehalten, falls sie nicht der konventionellen Meinung zu einem Thema entsprechen, oder falls man etwas

grundsätzlich Neues hat. Eine Technik zur Validation ist die Wiederholung, bei der man bestimmte Ideen Schritt für Schritt in Gedanken wiederholt oder praktisch durchführt, um zu prüfen, ob sie sinnvoll sind. Dies kann dann die Grundlage für weitere Reflexion sein.

Bei der *Aneignung* (im Orig. „appropriation“) wird das neugelernte Wissen unter Umständen tiefer in die eigene Persönlichkeit integriert (vgl. im weiteren Abschnitt Boud, Keogh und Walker 1985, 26–36). Dazu gehört neues Wissen, was das eigene Selbst betrifft und in die eigenen Wertvorstellungen übernommen wird. Bei dieser Art des Lernens können daher starke zusätzliche Gefühle auftauchen, die man mit dem Lernen assoziiert. Man betrachtet das Wissen als sein eigenes und neues Wissen kann vor dem Hintergrund dessen eingeordnet werden. Dieser Aspekt tritt nicht bei allem, was man lernt, auf.

Boud et al. sprechen ebenfalls davon, dass zu jedem Schritt neue Einsichten plötzlich und ungeplant auftauchen können und man sich dann erst diesen widmet, bevor man in den vorherigen Gedanken weitergeht (vgl. für diesen Abschnitt Boud, Keogh und Walker 1985, 26–36). Diese Einsichten sind dann der neue Gegenstand der Reflexion. Außer bei diesen spontanen Einsichten, empfehlen Boud et al., dass man sich während der Reflexion an den Ablauf dieser drei Etappen hält.

Diese vier Aspekte sind alle Teil der Neu-Evaluierung einer Erfahrung (vgl. im weiteren Abschnitt Boud, Keogh und Walker 1985, 26–36). Das Ergebnis dessen ist in Abbildung 2-2 (rechts) dargestellt. Das Ergebnis der Reflexion kann eine neue Art sein eine Tätigkeit durchzuführen, oder die Klärung eines Problems, oder die Entwicklung einer neuen Fähigkeit, oder ein neues Set an Ideen. Es kann sich das eigene Verhalten gegenüber ähnlichen Situationen in der Zukunft ändern, bspw. indem man in bestimmten Situationen selbstbewusster agiert oder eine neue Herangehensweise zeigt. Da die Ergebnisse der Reflexion verloren gehen können, wenn sie nicht an eine Aktion gebunden werden, ist es wichtig, dass der Lernende sich dazu bekennt (im Orig. „*commitment*“) das neue Wissen in der Praxis einzusetzen. Das Einsetzen des Wissens beendet dann die Reflexion, kann aber eine neue Reflexion auf Basis der Erfahrung des Einsetzens auslösen. Boud et al. sagen auch, dass es schwer ist, die neuen Ideen in die eigene Praxis zu überführen.

An der Stelle zeigt sich eine Trennung von Nachdenken und Handeln, was ähnlich der Arbeit von Kolb ist (siehe Abschnitt 2.4.3), aber dem *reflection-in-action* von Schön widerspricht (siehe Abschnitt 2.4.2).

Die Elemente, die bisher beschrieben wurden und in Abbildung 2-2 dargestellt sind, haben keinen zwingenden linearen Ablauf und sind nicht voneinander abhängig (Boud, Keogh und Walker 1985, 26–36). Manche der Aspekte, wie bspw. die Aneignung neuen

Wissens (s. o.), tauchen nicht in jedem Reflexionsprozess auf. Lernende wechseln fließend zwischen den einzelnen Schritten im Reflexionsprozess und gehen zu Erfahrungen zurück, um neue Details zu integrieren. Boud et al. geben zu bedenken, dass man im Falle von Schwierigkeiten im Reflexionsprozess diese Schritte und Aspekte als Hilfe zurate ziehen kann.

Zusammenfassend beschreiben Boud et al. einen detaillierten Reflexionsprozess, der darin besteht, die konkrete Erfahrung Revue passieren zu lassen und dann neu zu bewerten, um Ideen für die Zukunft abzuleiten und dann zu beabsichtigen diese in die Praxis umzusetzen. Boud et al. konzentrieren sich auf einen alleinigen Lerner, der ggf. von einem Lehrer in seiner Lernsituation unterstützt wird und beschreiben keine gruppendynamischen Prozesse. Im Wesentlichen beschränkten sich ihre Hinweise darauf, dass jemand in der Reflexion unterstützen kann. Gemeinsame Reflexion ist nicht erwähnt. Im Vergleich zu Kolb (siehe Abschnitt 2.4.3) und Schön (siehe Abschnitt 2.4.2) zeigt sich, dass der Reflexionsprozess wesentlich detaillierter ist. Während Boud, Keogh und Walker die Reflexion eher losgelöst von Handlungen sehen, ist dies bei Schön (siehe Abschnitt 2.4.2) mit dem Konzept der *reflection-in-action* wesentlich enger verknüpft.

#### **2.4.5 Weitere Aspekte von Reflexion und aktuellere Definitionen**

Nach dem in den vorherigen Abschnitten die Arbeiten unterschiedlicher Autoren näher beleuchtet wurden, werden in diesem Abschnitt jüngere Arbeiten zu Reflexion thematisiert.

Eine etwas ältere Beschreibung von Reflexion von Boyd und Fales beinhaltet ebenfalls ähnlich zu Boud et al. (siehe Abschnitt 2.4.4) die Sequenz, dass ein Problem untersucht wird, dass durch eine Erfahrung verursacht und in einer veränderten Perspektive resultiert (Boyd und Fales 1983). In einer Definition beschreiben Boyd und Fales Reflexion als Prozess, der eine Erfahrung aufklärt und eine veränderte Sichtweise das Ergebnis dieser Analyse einer Erfahrung ist (Boyd und Fales 1983). Als Auslöser wird häufig das Feststecken in einer Situation bezeichnet, wobei Reflexion durch allgemein unerwartete Aspekte ausgelöst werden kann. Dementsprechend können dies sowohl negative als positive Erfahrungen sein, wobei über Letztere seltener reflektiert wird (Gustafsson und Fagerberg 2004). Der Erfahrungsstand einer Person scheint Einfluss auf den Gegenstand der Reflexion zu haben, da Novizen und Experten über unterschiedliche Dinge reflektieren (Gustafsson und Fagerberg 2004).

Neben dem Modell von Boud et al. (siehe Abschnitt 2.4.4) gibt es noch andere Modelle, bei denen Reflexion in unterschiedliche Phasen eingeteilt ist. Das Modell von Atkins und

Murphy ähnelt dem von Boud et al. sehr stark, da drei Phasen vorhanden sind: 1) Auslöser ist das Erleben von unangenehmen Gefühlen oder Gedanken, 2) Analyse der Situation und 3) Entwickeln von neuen Perspektiven hinsichtlich der Situation (Atkins und Murphy 1993). Die letzte Phase wird von den Autoren als Lernen bezeichnet (Atkins und Murphy 1993). Bei Daudelin werden vier Phasen genannt: Benennung eines Problems, Analyse des Problems, Formulieren und Testen einer Theorie, um das Problem zu lösen, und die Aktion bzw. die Entscheidung, ob man agiert (Daudelin 1996). Damit liegt das Modell sehr dicht bei dem Reflexionsmodell von Kolb (siehe Abschnitt 2.4.3) und Boud et al. (siehe Abschnitt 2.4.4), die bei Daudelin zwar nicht direkt zitiert, aber in der relevanten Literatur genannt werden.

Høyrup sieht nicht nur das Feststecken in Tätigkeiten als Auslöser für Reflexion, sondern das Aufschieben von anderen Tätigkeiten, um Freiräume zum Reflektieren zu schaffen (Høyrup 2004). Ein weiterer Auslöser für Reflexion können eigene Gefühle sein (Sinclair und Tse 2008). Trotz des oft erwähnten Ablaufes, stellt ein Teilnehmer in einer Studie fest, dass Reflexion nicht immer einem Muster oder einer bestimmten Agenda folgt (Glazer, Abbott und Harris 2004, 36). Dies deutet darauf hin, dass der Prozess von Reflexion in der Praxis komplexer sein kann als in den Modellen angenommen. Dazu passend kritisieren manche Autoren die Modelle von Kolb (siehe 2.4.3) hinsichtlich der Pfade zum reflexiven Lernen als stark vereinfacht. Es wird empfohlen, weniger das Reflektieren über Daten anzuregen, als vielmehr das Reflektieren über Erfahrungen (Slovák, Frauenberger und Fitzpatrick 2017).

Durch die Lösung von Problemen kann Reflexion als Prozess verstanden werden, der eine zusammenhangslose Situation aufklärt und Sinn gibt (Clarà 2015). Dies wird so ähnlich bei Dewey beschrieben (siehe Abschnitt 2.4.1). Der Prozess von Reflexion hilft implizites Wissen in einen Kontext einzubetten bzw. zu erläutern (Järvinen und Poikela 2001). Hatton und Smith fügen hinzu, dass Reflexion absichtlich (im Orig. *deliberate*) stattfindet, mit der Zielsetzung der Verbesserung einer Handlung (Hatton und Smith 1995).

Man reflektiert allerdings nicht immer nur über die Erfahrung selbst, sondern darüber wie gut man eine Erfahrung erfassen bzw. das Problem betrachten kann (Marcu, Dey und Kiesler 2014). Reflexion wird als spontan und gewöhnlich betrachtet (Clarà 2015).

Eine Übersicht unterschiedlicher Definitionen von Reflexion über Zeit findet sich bei Nguyen et al. (2014).

#### 2.4.6 Ebenen von Reflexion

Häufig wird bei Reflexion festgestellt, dass diese eine unterschiedliche Qualität haben kann, womit gemeint ist, ob die Reflexion eher oberflächlich abläuft oder tiefer geht (Tigelaar u. a. 2008). Mit oberflächlicher Reflexion sind Fragen gemeint, wie man eine Situation lösen könnte und tiefergehende Reflexion impliziert moralische, politische oder emotionale Aspekte (Tigelaar u. a. 2008). Es gibt Modelle, welche den Grad der Reflexivität in unterschiedliche Ebenen unterteilen. Diese Modelle können bspw. helfen, Unterstützung für Reflexion zu strukturieren und zu entwickeln (siehe Abschnitt 2.8).

Ein Modell von Fleck und Fitzpatrick unterteilt Reflexion in fünf unterschiedliche Ebenen (im Orig. *Level*), die im Folgenden kurz beschrieben werden (Vgl. im Folgenden Fleck und Fitzpatrick 2010). Die Ebenen sind hierarchisch angeordnet und eine höhere Ebene impliziert eine intensivere Reflexion:

0. *Description (Revisiting)*: Dies umfasst einfache Beschreibungen einer Situation ohne Interpretation oder Erklärungen und wird als nicht reflexiv bezeichnet.
1. *Reflective Description (Revisiting with Explanation)*: Auf dieser Ebene verfügen die Erklärungen bereits über eigene Interpretationen, warum etwas passiert ist. Allerdings ist auf dieser Ebene nur eine oberflächliche Analyse der Situation vorhanden bzw. es werden zur Analyse keine unterschiedlichen Perspektiven eingenommen.
2. *Dialogic Reflection (Exploring Relationships)*: Auf dieser Ebene gibt es bereits mehrere Zyklen, in denen sich der Reflektierende mit dem Problem auseinandergesetzt hat. Es gibt mehrere Blickwinkel und Erklärungsversuche für eine Situation. Es werden ebenfalls Zusammenhänge gesucht.
3. *Transformative Reflection (Fundamental Change)*: Auf dieser Ebene wird beim Reflektieren einer Aktion oder Situation versucht persönliche Sichtweisen auf einen Sachverhalt zu hinterfragen. Die Zielsetzung ist die Veränderung der eigenen Praxis bzw. dem eigenen Verständnis des Sachverhalts.
4. *Critical Reflection (Wider Implications)*: Die höchste Ebene der Reflexion bezieht sich auf das Einbeziehen breiterer Themenkomplexe, wie bspw. soziale oder ethische Sachverhalte.

Hinsichtlich der Unterstützung der Ebenen der Reflexion, wird hervorgehoben, dass die Unterstützung, bspw. die Technologie, lediglich eine Gelegenheit bieten kann, aber dass die Reflexion selbst immer noch von den Personen durchgeführt werden muss (Fleck und



Fitzpatrick 2010). Häufig wird bei der Reflexion nur eine der unteren Ebenen erreicht. In einer Studie von Hatton und Smith haben 70 % der Teilnehmer eine Ebene beschreibender Reflexion, was der zweiten Ebene entspricht, erreicht (Hatton und Smith 1995). Um effektiv zu sein, muss bei der Reflexion aber nicht die höchste Ebene erreicht werden (Fleck und Fitzpatrick 2010; Hatton und Smith 1995).

Moon unterscheidet zwischen informeller Reflexion und akademischer Reflexion (im Orig. *academic reflection*) (Vgl. für diesen Abschnitt Moon 2006a, 135–137). Unter akademischer Reflexion beschreibt sie einen strukturierten und formelleren Prozess, in dem eher thematisiert wird, was die Reflexion als Ziel hat. Zusätzlich findet diese Form der Reflexion eher in einem akademischen Rahmen statt und wird von Mentoren begleitet. Dieser Begriff wird von informeller Reflexion bzw. *common-sense reflection* abgegrenzt, bei der Reflexion eher informell durchgeführt wird. Walsh interpretiert die Verwendung des Reflexionsbegriffs bei Boud et al. (2006) als eher informell (Walsh 2009).

Zusammengefasst lässt sich Reflexion in unterschiedliche Ebenen einordnen, die den Grad der Tiefe, mit der sich jemand mit einem Thema auseinandersetzt, abbildet. Darüber leiten bspw. Fleck und Fitzpatrick unterschiedliche Möglichkeiten der Unterstützung ab, welche in Abschnitt 2.8 diskutiert werden (Fleck und Fitzpatrick 2010).

## **2.5 Kollaborative Reflexion**

Kemmis spricht im gleichen Buch wie Boud et al. davon, dass Reflexion ein inhärent sozialer Prozess ist und kein rein individueller Prozess, da man die Welt beeinflusst, sobald man auf Basis der Reflexionsergebnisse handelt (Kemmis 1985, 140–149). Dennoch fokussieren sich Schön (siehe Abschnitt 2.4.2; Schön 1983), Kolb (siehe Abschnitt 2.4.3; D. A. Kolb 1984) und Boud et al. (siehe Abschnitt 2.4.4; Boud, Keogh, and Walker 1985) auf individuelle Reflexion und vernachlässigen in ihren Modellen eine Ebene, in der mehrere Personen gemeinsam auf Augenhöhe reflektieren. Dies war Anlass für weitere Forschung in diese Richtung (bspw. Prilla et al. 2011; Kim and Lee 2002). Die soziale Komponente der Reflexion beschränkt sich bei Schön, Kolb und Boud auf einem Mentor–Mentee- bzw. Lehrer–Schüler-Prinzip, bei dem die Lernenden zusammen mit den Betreuenden reflektieren. Dies hatte oft nur eine Dimension, in der einer dem anderen Aspekte aufzeigt, die dieser übersehen hat (Cressey, Boud und Docherty 2006). Eine Reflexion von Personen auf gleicher Augenhöhe, sprich auf einem gleichen Ausbildungsstand, wird nicht umfassend beschrieben bzw. in den Modellen berücksichtigt. Die Forschung hat kollaborative Reflexion nicht näher betrachtet, wobei

es dennoch bereits breit in der Praxis eingesetzt bzw. angewandt wurde, bspw. im Gesundheitssektor (Gustafsson und Fagerberg 2004; Kim und Lee 2002).

Es gibt einerseits den Begriff der kollaborativen Reflexion, der aktuell in vielen Publikationen genutzt wird (bspw. Renner u. a. 2019; Prilla und Herrmann 2017). Es gibt zusätzlich noch den Begriff der kollektiven Reflexion, der von der Bedeutung her synonym verwendet wird (Cobb u. a. 1997; Nissilä 2005; Levina 2005; Cressey, Boud und Docherty 2006). In dieser Dissertation wird im Folgenden nur der Begriff der kollaborativen Reflexion verwendet, der insbesondere die Zusammenarbeit unterstreicht.

### 2.5.1 Begriffsklärung

Prilla et al. definieren kollaborative Reflexion als sozialen Reflexionsprozess, der durch Kommunikation zwischen Personen aktiviert wird, die ihre eigenen Erfahrungen zum Reflexionsprozess beisteuern, mit der Zielsetzung, um gemeinsam daraus zu lernen (Prilla, Herrmann und Degeling 2013; Prilla, Degeling und Herrmann 2012; Høyrup 2004). Weiterhin führen Prilla et al. aus, dass diese Reflexion auch spontan auftauchen kann und keine feste Gruppe benötigt. Dies entspricht eher dem Begriff der informellen Reflexion (siehe Abschnitt 2.4.6). Dennoch findet bei kollaborativer Reflexion der eigentliche Reflexionsprozess immer noch bei einem selbst statt (Daudelin 1996), wenn auch in einem sozialen Kontext (Høyrup 2004). Eine Gruppe kann die Erfahrungen ihrer Mitglieder zusammenfassen, um gemeinsam darüber zu reflektieren (Høyrup 2004), und gemeinsam daraus zu lernen (Dyke 2006). Bei einer kollaborativen Reflexion mit anderen ist man anderen Ansichten ausgesetzt und es können so andere Weltanschauungen mit einbezogen werden, die aber nicht von jedem übernommen werden müssen (Dyke 2006; Kim und Lee 2002). Dyke unterscheidet weiterhin zwischen *primären* Erfahrungen, welche man selbst gemacht hat, und *sekundären* Erfahrungen, welche andere gemacht haben. Er fügt hinzu, dass es Strömungen gibt, bei denen primäre Erfahrungen bevorzugt und sekundäre Erfahrungen ablehnend betrachtet werden. Bei kollaborativer Reflexion kann man von beiden Erfahrungsarten lernen. Auf diese Weise lässt sich der eigene Horizont bei kollaborativer Reflexion erweitern. Prilla et al. differenzieren kollaborative Reflexion zu individueller Reflexion insbesondere dahingehend, dass bei kollaborativer Reflexion Kommunikations- und Koordinationsprozesse hinzukommen (Prilla, Degeling und Herrmann 2012). Der Gegenstand der kollaborativen Reflexion kann eine eigene oder eine gemeinsame Tätigkeit sein (Prilla, Degeling und Herrmann 2012).

Dillenbourg unterstreicht, dass die Kollaboration bzw. das Vorhandensein einer Gruppe an sich nicht ausreichend ist, um zu lernen und kontrastiert, dass man nicht lernt, indem

man alleine ist (Dillenbourg 1999). Dementsprechend sieht Dillenbourg Aktivitäten, die durchgeführt werden müssen, damit man in einer Gruppe lernen kann. Dazu passend beschreibt Boud, dass die Erfahrung die Grundlage zum Lernen ist, aber alleine nicht ausreicht und es einen Prozess gibt, der die Erfahrung in Wissen transformiert (Dyke 2006). Prilla et al. geben zu bedenken, dass nicht jede Interaktion, in der man gemeinsam ein Problem diskutiert, kollaborative Reflexion ist (Prilla, Herrmann und Degeling 2013).

Kollaborative Reflexion wird im Vergleich zu individueller Reflexion als potenter beschrieben, da man durch die Einsichten anderer profitieren kann oder zu eigenen Problemen Hilfe bekommen kann (Kim und Lee 2002; Daudelin 1996; Høyrup 2004). Die beiden Formen individuelle Reflexion und kollaborative Reflexion sind nicht voneinander getrennt und so kann individuell erworbenes Wissen in der Gruppe gemeinsam erweitert werden und umgekehrt (Kim und Lee 2002).

Høyrup differenziert zwischen drei unterschiedlichen Ebenen der Reflexion: eine Ebene der individuellen Reflexion, eine Interaktionsebene und eine organisationale Ebene. Die individuelle Ebene zieht er aus der Literatur von Dewey (siehe Abschnitt 2.4.1; Dewey 1933), Schön (siehe Abschnitt 2.4.2; Schön 1983), und Boud et al. (siehe Abschnitt 2.4.4; Boud, Keogh, and Walker 1985). Für die Interaktionsebene führt er Raelin als Quelle an (Raelin 2002), der eine soziale Ebene in der Reflexion skizziert, bei der mehrere Personen zusammen reflektieren. Die organisationale Ebene beschreibt das Lernen im beruflichen Kontext (siehe Abschnitt 2.6). Dazu gehört bspw. das Hinterfragen von Annahmen, welches Høyrup als kritisch ansieht, da es einerseits Struktur und Rahmenbedingungen im Unternehmen setzt, aber andererseits wichtig ist, um aktuell allgemeinen Herausforderungen zu begegnen (siehe Abschnitt 2.7.2 - Critical Reflection; (Høyrup 2004)). Järvinen und Poikela beschreiben ein ähnliches Modell bestehend aus drei Ebenen: lernen auf individueller Ebene, auf Gruppenebene und auf organisationaler Ebene (Järvinen und Poikela 2001).

Gelegenheiten für kollaborative Reflexion sind beispielsweise gemeinsame Projekte, bei denen man gegenseitig Feedback geben oder sich bei Lernstrategien helfen kann (Kim und Lee 2002). Kollaborative Reflexion hat gezeigt, dass bspw. Lehrer nicht nur über Fragen von Lehransätzen diskutieren, sondern über politische oder emotionale Herausforderungen im Beruf (Tigelaar u. a. 2008). Untersuchungen bei Managern haben gezeigt, dass geringe Zeitaufwände in kollaborative Reflexion helfen über eine Situation mehr zu lernen (Daudelin 1996).

### 2.5.2 Elemente kollaborativer Reflexion

In der bisherigen Beschreibung von kollaborativer Reflexion wurde nur die Bedeutung gemeinsamer Reflexion hervorgehoben, und dass man voneinander lernen kann. In diesem Abschnitt wird konkretisiert, wie dieses gemeinsame Lernen funktioniert bzw. welche Voraussetzungen erfüllt sein müssen, um Lernen zu ermöglichen.

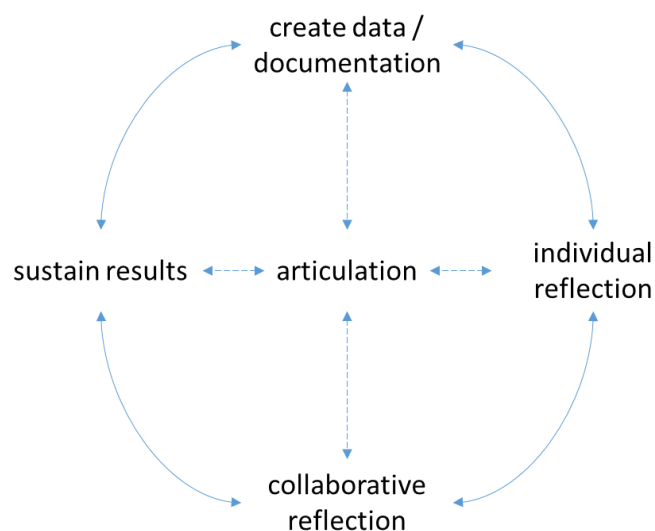
Mehrere Autoren heben hervor, dass der wichtigste Bestandteil kollaborativer Reflexion das Teilen der eigenen Erfahrung bzw. die Artikulation der eigenen Erfahrung ist bzw. das Explizieren eigenen Wissens oder Dingen, die unklar sind (Daudelin 1996; Prilla, Herrmann und Degeling 2013; Høyrup 2004; Järvinen und Poikela 2001; Scott 2010; Kim und Lee 2002; Levina 2005). Kollaborative Reflexion kann auf diese Weise helfen gemeinsame Ziele zu erreichen (Kim und Lee 2002). Der Vorteil ist, dass jeder Teilnehmer gefordert ist seine Sichtweise zu argumentieren und so zu versuchen seine Reflexionspartner zu überzeugen (Fleck und Fitzpatrick 2006; Dillenbourg 1999). Das Lernen tritt dann auf, wenn jemand Verknüpfungen zwischen eigenem Wissen und dem Wissen anderer herstellt bzw. wenn eine Gruppe ihre Sichtweisen zusammenfügt und gemeinsam darüber reflektiert (Daudelin 1996; Høyrup 2004; Prilla, Herrmann und Degeling 2013; Glazer, Abbott und Harris 2004). So kann eine Gruppe zusammen Theorien entwickeln, die dann wiederum gemeinsam oder allein in Richtung Umsetzung gebracht werden. Ein weiterer Punkt ist der Austausch von Lösungsvorschlägen auf Basis von eigenen Erfahrungen (Boerboom u. a. 2011).

Van Woerkom und Croon detaillieren die genannte Artikulation eigener Erfahrungen noch weiter und nennen unterschiedliche Aspekte kollaborativer Reflexion (vgl. im Folgenden Woerkom und Croon 2008):

- „*Openness about mistakes*“: Da Fehler eine sehr gute Möglichkeit zu Lernen bieten (Keith und Frese 2005), ist die Empfehlung mit Fehlern offen umzugehen und als Ausgangspunkt zur Reflexion zu nutzen.
- „*Asking for feedback*“: Feedback bekommen hilft sowohl beim Verständnis als bei der persönlichen Motivation. Andere Studien verdeutlichen ebenfalls die Stärken von Feedback (Kim und Lee 2002) bzw. nennen Feedback als Ausgangspunkt für Reflexion (Daudelin 1996; Schön 1983).
- „*Critical opinion-sharing*“: Hier unterscheiden van Woerkom und Croon Reflexion von *critical reflection*, bei der man nicht mehr nur über das *wie* der Problemlösung nachdenkt, sondern auch über das *warum* und Annahmen infrage stellt (siehe Abschnitt 2.7.2).

- „*Challenging groupthink*“: Bei kollaborativer Reflexion sind die Teilnehmer nicht immer einer Meinung, und es gibt auch Situationen, in denen ein Teilnehmer der Meinung der Gruppe aktiv widerspricht. Dies bietet weitere Situationen, um zu lernen.

Bei der kollaborativen Reflexion diskutiert man über einen bestimmten Sachverhalt. Es bietet sich an Artefakte heranzuziehen, um bestimmte Sachverhalte leichter erklären zu können, die ähnlich wie Schulungsmaterial in formalen Trainings genutzt werden (Prilla u. a. 2011). In mehreren Studien in Krankenhäusern haben Prilla et al. beobachtet, dass dort Patientenakten als Artefakte genutzt wurden, um während einer Übergabe zwischen Krankenpflegern in unterschiedlichen Schichten die Reflexion zu unterstützen (Prilla, Herrmann und Degeling 2013). Es können Bilder genutzt werden, um die eigene Argumentation zu stützen, bzw. eine Aussage zu verdeutlichen (Fleck und Fitzpatrick 2006). Weitere Ideen für Artefakte zur Reflexionsunterstützung sind persönliche Notizen über Vorgänge (Hatton und Smith 1995), oder Workflowdaten (Prilla u. a. 2011; Lin u. a. 1999). Um die Reflexion zwischen Managern zu unterstützen, hat Daudelin ein *Community Reflection Workbook* erstellt, in dem Anleitungen sind, welche die Funktionsweise des Reflexionsprozesses erklären, inklusive mehrerer Fragen, die eine Reflexion anregen sollen (Daudelin 1996). In dem Workbook wurden so Phasen individueller und kollaborativer Reflexion verbunden. Prilla et al. skizzieren, dass Protokolle aus Sitzungen als Artefakte für Reflexion zu nutzen sind, und regen an, alle Notizen, die Erfahrungen artikulieren, als mögliche Artefakte zur Unterstützung von Reflexion aufzufassen (Prilla, Herrmann und Degeling 2013).



**Abbildung 2-3: Kollaborative Reflexion und die Rolle von Artikulation**  
(Prilla, Degeling und Herrmann 2012, Figure 3; Prilla und Renner 2014, Figure 1)

Prilla et al. beschreiben ein Modell kollaborativer Reflexion (siehe Abbildung 2-3) in dem der Startpunkt das Festhalten einer Erfahrung ist, was eine Dokumentation oder das Erfassen von Daten beinhalten kann (vgl. im Folgenden Prilla, Degeling und Herrmann 2012; Prilla und Renner 2014). Dem folgt eine Phase individueller Reflexion, in der beispielsweise der Bedarf nach kollaborativer Reflexion auftauchen kann. Bei erfolgreicher kollaborativer Reflexion kommt eine Phase, in der die Ergebnisse dokumentiert werden. Dennoch ist wichtig zu unterstreichen, dass es sich nur um Reflexion handelt, wenn eine Handlungsabsicht dahintersteht. Die *articulation* verbindet alle Phasen, da der Bedarf entstehen kann, bestimmte Erlebnisse anderen mitzuteilen oder selbst zu verschriftlichen. Zwischen allen Phasen ist eine Rückwärtsbewegung möglich, um zu verdeutlichen, dass man fließend zwischen den Phasen wechseln kann, sollte aus einer vorherigen Phase bspw. weiteres Wissen fehlen.

In dieser Arbeit wird Reflexion in Anlehnung an die Beschreibung von Schön als das Zurückkehren zu Erlebnissen und deren systematische Neubewertung hinsichtlich des aktuellen Erfahrungsschatzes betrachtet. Diese Neubewertung zielt darauf ab, künftige Handlungsweisen für die Zukunft zu identifizieren und zu planen. Die Reflexion kann dabei allein oder im Austausch mit mehreren Personen erfolgen.

### **2.5.3 Gelegenheiten für kollaborative Reflexion**

Um kollaborative Reflexion zu unterstützen, müssen Gelegenheiten geschaffen werden. Es eignen sich einerseits formale Sitzungen in der Lehre, wie die Beispiele bei Schön und Boud et al. zeigen (Schön 1983; Boud, Cressey und Docherty 2006; Clouder 2000; Shields 1995). Teams, die sich speziell zum Reflektieren treffen, sind eine weitere Gelegenheit (Høyrup 2004; Raelin 2002). Zusätzlich eignen sich im beruflichen Kontext geplante Sitzungen, wie Projektmeetings oder Teammeetings, als Gelegenheit zur Reflexion (Daudelin 1996). Andererseits lässt sich kollaborative Reflexion ungeplant in informellen Gesprächen mit Kollegen oder Freunden durchführen (Daudelin 1996).

In einer Studie zur Ausbildung von Physiotherapeuten wurde evaluiert, ob ein Modell von 2 Studenten pro Praxisausbilder funktioniert, da es in der Vergangenheit immer nur ein 1:1 Verhältnis in der Betreuung gab (vgl. für diesen Abschnitt Morris und Stew 2007). Es wurde gezeigt, dass die Studenten mehr Gelegenheiten zur Reflexion haben, da sie untereinander reflektieren können und nicht warten müssen bis ihr Betreuer verfügbar ist. Es wurde beobachtet, dass Studenten ihre Patienten mit in die Reflexion einbezogen haben, um ein umfangreicheres Bild der Situation zu bekommen.

#### 2.5.4 Der Arbeitsplatz als Reflexionsumgebung

Im beruflichen Kontext reflektieren wirkt sich auf zwei Ebenen aus: Zum einen reflektiert man über die zu erledigenden bzw. erledigten Aufgaben und zum anderen sind Kommunikations- und Koordinationsprozesse für die Reflexion notwendig (Prilla u. a. 2011). Die Reflexion kann sowohl in-action als auch on-action sein (Prilla, Degeling und Herrmann 2012). Clouder fügt hinzu, dass Reflexion intuitiv sein muss, da es sonst in Zeiten von viel Druck ineffizient ist (Clouder 2000).

Unterschiedliche Formen, um kollaborative Reflexion in Unternehmen zu organisieren bzw. das organisationale Lernen zu unterstützen, finden sich bei Vince (vgl. für diesen Abschnitt Vince 2002). Dieser nennt vier Formen: *peer consultancy groups*, *organisational role analysis*, *communities of practice*, und *'group relations' conferences*. Mit *peer consultancy groups* meint er Gruppen von drei Managern, die sich gegenseitig bei der Problemlösung unterstützen, wobei einer ein Problem präsentiert, der zweite bei der Problemlösung unterstützt und der dritte den Prozess der Problemlösung beobachtet. Bei der *organisational role analysis* reflektieren zwei Personen zusammen, wie sich ihre Rolle im Unternehmen und beim Wandel der Organisation verhält. Bei *communities of practice* beschreibt er intraorganisationale Gruppen mit gleichen Tätigkeiten, die sich in der Organisation engagieren wollen (siehe Abschnitt 2.1). Die *'group relations' conferences* beschreiben eine Gelegenheit, in der mehrere Personen über die Beziehung zwischen Autorität, Führung und Transformation reflektieren. Vince unterstreicht aber, dass die Reflexion an sich weiterhin unter der Kontrolle der Mitarbeiter sein muss. All diese Formen bieten sich aus organisationaler Sicht an, um das Lernen in der Organisation außerhalb formaler Trainings zu stärken.

Prilla et al. skizzieren unterschiedliche Rollen, die an einer kollaborativen Reflexion beteiligt sind (vgl. für diesen Abschnitt Prilla, Herrmann und Degeling 2013): Zum einen gibt es einen *reflection initiator*, der das Thema initial aufbringt und dazu angeregt über das Thema zu reflektieren. Der Rolle *topic owner* gehören Personen an, die eine spezifische Reflexionssitzung auslösen, aber auch solche, die ein Thema in Vertretung für den reflection initiator starten. Der *reflection sparring partner* ist jemand, der spontan zu einer Reflexion hinzugezogen werden kann, um Fragen zu klären. Diese Rolle taucht nicht bei jeder Reflexion auf und die Reflexion kann danach weiter individuell verlaufen. Die Rolle *reflection executive* ist vorhanden, wenn in einer Gruppe, die zusammen über ein Problem reflektiert hat, wenige Personen danach die Entscheidungen treffen, wie im weiteren Verlauf vorgegangen werden soll. Zusätzlich gibt es *reflection participants*, die in einer kollaborativen Reflexion eigene Erfahrungen einbringen und mit über das Problem diskutieren, was auch Daudelin als Aufgabe bestätigt (Daudelin 1996).

Während der Arbeit reflektiert man überwiegend mit Kollegen oder mit Vorgesetzten (Kim und Lee 2002), und im privaten Bereich sind oft Lebensabschnittsgefährten, Freunde, Therapeuten oder Kirchenvertreter Partner in der Reflexion (Daudelin 1996). Prilla et al. haben beobachtet, dass Reflexion nur unter Personen mit der gleichen beruflichen Ausrichtung stattfindet (Prilla, Herrmann und Degeling 2013). Dies deckt sich mit den Beobachtungen von Morris und Stew (Morris und Stew 2007). Es deckt sich aber nicht vollständig mit den Szenarien bei Schön, Kolb und Boud et al., in denen oft eine Reflexion von Lehrern und Schüler beschrieben wird (siehe Abschnitt 2.4). Der Grund kann darin begründet sein, dass man häufig über Tätigkeiten reflektiert, in denen der Ablauf oder Erfolg nicht wie gewünscht war, und man über solche Aspekte ungern mit Vorgesetzten oder gegenüber Kunden/Klienten spricht (Clouder 2000). Ähnlich kann es sein, dass man über solche Aspekte nicht mit Novizen spricht, da diese ggf. noch nicht den Erfahrungsstand haben, um einem helfen zu können. Dillenbourg sieht das allgemeiner, und sagt, dass Zusammenarbeit häufig unter Personen stattfindet, sobald eine Symmetrie zwischen den Personen gegeben ist bspw. bei ähnlichem Wissensstand oder ähnlichem Status (Dillenbourg 1999). Unter Schülern ist es aus Sicht der Dozenten wichtig, dass diese gut miteinander auskommen, um den Erfolg kollaborativer Reflexion zu ermöglichen (Morris und Stew 2007). Dillenbourg unterstreicht, dass der Wissensstand bei zwei Personen nie gleich ist. Cressey et al. fügen hinzu, dass auch gemeinsame Ziele ein Grund für gemeinsame Reflexion sein kann (Cressey, Boud und Docherty 2006).

Bei Daudelin gibt eine *Peer-Group*, in der nur Personen sind, die mit der reflektierenden Person in keinem Hierarchieverhältnis stehen (Daudelin 1996). Ausgehend von Daudelins Beschreibung wird impliziert, dass diese fachfremd sind bzw. sich nicht gut in dem Fachgebiet der reflektierenden Personen auskennen. Daudelin benennt drei Gründe, warum in der Gruppe weniger gelernt wurde bzw. dieses beobachtet wurde: Die Personen sind übermäßig lange mit Erfahrungsaustausch bzw. mit der Suche von vergleichbaren Aspekten in ihren Erfahrungen beschäftigt; aufgrund der Vielzahl ausgetauschter Erfahrungen wurden einzelne Erfahrungen nicht in der Tiefe diskutiert; und die Gruppe hat erklärendes Material zum Ablauf von Reflexion und reflexive Fragen nicht genutzt. Hier lässt sich zusätzlich entnehmen, dass Erfahrungen durchaus in der Tiefe diskutiert werden müssen und nicht nur in der Breite. Gleichzeitig haben die Personen in der *Peer-Group* mehr über andere gelernt als über sich selbst, was verdeutlicht, dass es sinnvoll ist, mit Personen gleicher beruflicher Ausrichtung bzw. Spezialisierung zu diskutieren. Eine Ausnahme ist die Zielsetzung, einen Überblick über das Lernen einer Gruppe zu bekommen (Daudelin 1996).



## 2.6 Einsatz von Reflexion aus der Sicht eines Unternehmens

Wie in den vorherigen Abschnitten beschrieben, eignet sich Reflexion als Mechanismus für Lernen (siehe Abschnitte 2.4 und 2.5). Im beruflichen Kontext lässt sich feststellen, dass Mitarbeiter bis zu 70 % der Zeit für Lernen mit informellen Lernen verbringen, anstatt in formellen Umgebungen wie Trainings oder Schulungen (Clardy 2018). Somit ist Lernen durch Erfahrung der Hauptteil, durch den Mitarbeiter am Arbeitsplatz dazulernen (Eraut 2004).

Dieser Abschnitt fokussiert sich allerdings weniger auf die Perspektive eines lernenden Mitarbeiters in Unternehmen und konzentriert sich mehr auf die Perspektive der Unternehmen selbst. Renner stellt fest, dass Reflexion sowohl von den Mitarbeitern (bottom-up) als auch von dem Unternehmen selbst ausgehen kann, sobald jemand in einer Management-Position aus Reflexion heraus Änderungen veranlasst (top-down) (Renner u. a. 2019).

Walsh stellt fest, dass aus Sicht der Unternehmen das Lernen und die damit verbundenen Aktivitäten Einzelner weniger wichtig sind und viel mehr das Endergebnis zählt (Walsh 2009). Reflexion wird als Kernprozess organisationalen Lernens gesehen (Høyrup 2004). Bei diesem Prozess profitiert neben den Einzelnen zusätzlich das Unternehmen selbst von der Reflexion der Mitglieder (Boud, Cressey und Docherty 2006, 19–20; Walsh 2009; Knipfer u. a. 2013). Unternehmen profitieren von der Reflexion der Mitarbeiter, da viel implizites Wissen (im Orig. *tacit knowledge*) bei der Reflexion sichtbar gemacht wird (Knipfer u. a. 2013; Eraut 2000). Zusätzlich profitiert ein Unternehmen durch Reflexion indirekt, da Mitarbeiter zunächst über eigene Probleme bzw. eigene Fähigkeiten reflektieren (Walsh 2009). Meetings sind ein wichtiger Baustein für Reflexion, da so Gelegenheiten geschaffen werden, in denen Erfahrungen ausgetauscht werden etc. (Høyrup 2004).

Vince beschreibt unterschiedliche Arten inwieweit (kollaborative) Reflexion als Prozess einem Unternehmen helfen kann (zur praktischen Umsetzung einiger Ideen, wie Mitarbeiter reflektieren können siehe Abschnitt 2.5.4) (vgl. für diesen Abschnitt Vince 2002). Zum einen unterstützt Reflexion organisationale Annahmen zu hinterfragen, bei der Machtverhältnisse (im Orig. *power relations*) offenbart werden. Andererseits kann Reflexion helfen resultierende Effekte, bspw. Unsicherheit, aus der Offenbarung der Machtverhältnisse abzudecken. Zusätzlich können Unternehmen Mitarbeitern helfen, von individueller zu kollaborativer Reflexion zu wechseln.

Um Mitarbeiter (kollaborative) Reflexion zu ermöglichen, muss ein Unternehmen sowohl Zeit als auch direkte Unterstützung einbringen (Bolton 2010).

Zusammengefasst lässt sich aus diesem Abschnitt entnehmen, dass (kollaborative) Reflexion nicht nur ein wichtiger Prozess für Einzelne, sondern auch ein Prozess ist, durch den Unternehmen profitieren können. Aus diesem Grund ist es wichtig, dass Unternehmen informelles Lernen der Mitarbeiter fördern.

## **2.7 Abgrenzung von anderen Konzepten**

In der Literatur tauchen andere Konzepte auf, die der Reflexion bzw. kollaborativen Reflexion ähnlich sind, aber doch Unterschiede aufweisen. Häufig werden die Begriffe synonym verwendet (D'Cruz, Gillingham und Melendez 2007). In diesem Abschnitt werden artverwandte Lernansätze von den zuvor beschriebenen Konzepten der Reflexion abgegrenzt, um das Konzept weiter zu schärfen.

### **2.7.1 Productive Reflection**

Cressey et al. bringen das Konzept der *productive reflection* auf, das mit organisationaler und individueller Reflexion verbunden ist (Cressey, Boud und Docherty 2006). Sie beschreiben, dass sowohl die reflektierenden Einzelpersonen als auch die Organisation von dem Prozess profitieren. Cressey et al. beziehen sich auf Reflexionsprozesse, die zwar ortsungebunden stattfinden können, aber immer im Kontext einer arbeitsbezogenen Tätigkeit stattfinden. Diese Tätigkeit bestimmt immer den Rahmen dessen, was reflektiert wird. Der Fokus liegt laut Cressey et al. bei der *productive reflection* immer auf der Aktion, die aus dem Ergebnis der Reflexion folgt, anstatt auf dem Verstehen eines Sachverhalts und die direkt im Arbeitskontext eingesetzt werden kann. Die *productive reflection* wird bei Cressey et al. mit einer sozialen Dimension unter Einbezug unterschiedlicher Stakeholder, bspw. mit Vorgesetzten oder Kunden, beschrieben. Somit grenzt sich das Konzept von der bisherigen Reflexion bzw. kollaborativen Reflexion dahingehend ab, dass ein klarer Fokus auf arbeitsbezogene Tätigkeiten liegen (und der private Kontext ausgeschlossen wird) und auch, dass die Aktion im Vordergrund steht.

### **2.7.2 Critical Reflection**

Bei manchen Autoren wird *critical reflection* zunächst ähnlich beschrieben, wie der Begriff der Reflexion, der bereits definiert wurde (siehe Abschnitt 2.4) (Reynolds 1999).

*Critical reflection* wird als Prozess, bei dem die reflektierende Person nicht über das Problem, sondern über den Problemlösungsansatz und die eigenen Annahmen reflektiert, beschrieben (Mezirow 1990; Reynolds 1999; Kemmis 1985; Brookfield 2010). Damit

wird eine tieferliegende Ebene unter der Problemlösung selbst skizziert, als das was bei Schön (siehe Abschnitt 2.4.2; Schön 1983), Kolb (siehe Abschnitt 2.4.3; Kolb 1984), und Boud et al. (siehe Abschnitt 2.4.4; Boud, Keogh, and Walker 1985) als Reflexion beschrieben wird. Reynolds hingegen bezieht critical reflection mehr thematisch auf das Reflektieren über soziale oder politische Prozesse, die über das eigentliche Problem oder die eigene Tätigkeit hinausgehen (Reynolds 1999). Diese Beschreibung passt ebenfalls zu dem vierten Level von Reflexion von Fleck und Fitzpatrick, welches dort als „*Critical Reflection: Wider Implications*“ bezeichnet wird, und moralische, ethische, sozio-historische und politisch-kulturelle Themen in die Reflexion mit einbezieht (Fleck und Fitzpatrick 2010).

Es ist festzustellen, dass diese tiefergehende Ebene der Reflexion in der Praxis seltener erreicht wird, als die übliche Reflexion über die eigene Tätigkeit (Fleck und Fitzpatrick 2010), bspw. beim Schreiben eines reflexiven Berichts (Wong 1985; Hatton und Smith 1995).

Es lässt sich zusammenfassen, dass sich Reflexion auf das Lösen eines konkreten Problems in der eigenen Tätigkeit bezieht und critical reflection sich mehr auf das Lösen über übergeordneten sozialen oder kulturellen Problemen bezieht (Brookfield 2010).

### **2.7.3 Critical Thinking**

Der Begriff *critical thinking* ist schwierig vom Begriff der Reflexion zu trennen, da bereits Dewey die Begriffe von *reflective thinking* und *critical thinking* häufig synonym verwendet hat (Dewey 1933; Scott 2010). Teilweise wird critical thinking mit critical reflection gleichgesetzt (Ghaye 2010). Andere Autoren haben diese Basis aufgegriffen, und beschreiben *critical thinking* als erfahrungsbasiertes Lernen (Shields 1995). Andere Definitionen des Begriffs beschreiben critical thinking als disziplinierten Ansatz Probleme zu lösen (Reynolds 1999).

Garrison stellt fest, dass critical thinking ein übergeordneter Begriff ist, und die Begriffe des *creative thinking* und Problemlösen mit einbezieht (Garrison 1991). Er sieht critical thinking als eher deduktiven Ansatz des Denkens, der auf schlecht definierte (im Orig. *ill-defined*) und abstrakte Probleme abzielt. Kreatives Denken hingegen sei mehr für die Entwicklung von Ideen. Ghaye hingegen sieht die Begriffe von creative und critical thinking als unterschiedlich und sich gegenseitig helfend (Ghaye 2010). Ghaye bezeichnet critical thinking als Prozess, in dem Ideen sowohl im Kontext künftigen Einsatzes bewertet werden, sondern auch als Prozess, in dem Ideen selektiert werden (Ghaye 2010).

Garrison sieht Unerwartetes als Auslöser von critical thinking, womit es inhaltlich dem Begriff der Reflexion ähnelt (Garrison 1991). Brookfield beschreibt eine Phase, in der ein auslösendes Ereignis am Anfang steht (Brookfield 1987). Bei anderen Autoren taucht die Erfahrung nicht als Komponente auf (Ennis 1993). Ähnlich der sozialen Ebene in kollaborativer Reflexion sieht Thayer-Bacon eine soziale Dimension in critical thinking (Thayer-Bacon 1993).

Critical thinking wird als korrektes Bewerten von Aussagen (im Orig. „[...] correct assessing of statements [...]“) bezeichnet (Ennis 1993). Obwohl die Begriffe von Reflexion und critical thinking sehr dicht beieinanderliegen, lässt sich feststellen, dass in der Literatur zu Reflexion vom Einbeziehen von Gefühlen gesprochen wird (siehe Abschnitt 2.4.4; Boud, Keogh, and Walker 1985), während bei critical thinking dies keine Erwähnung findet. Bei critical thinking hingegen, wird häufig von logischem Denken und Begründen (*logical reasoning*) gesprochen (Garrison 1991), was sich in dem disziplinierten Ansatz zur Problemlösung bei Reynolds wiederfindet (1999).

#### **2.7.4 Reflexivity und Team Reflexivity**

Oft wird zwischen den Begriffen der Reflexion und der reflexivity kein Unterschied gemacht (Neill 2006; Finlay 2002; Renner u. a. 2019). Freshwater und Rolfe skizzieren weiterhin drei Arten von Reflexivität: eine Art meta-reflection, bei der man über die Reflexion selbst reflektiert mit einem Bezug auf einen sozialen und politischen Kontext und Reflexion der eigenen Praxis (Freshwater und Rolfe 2001). Die zweite Art mit Bezug auf soziale und politische Kontexte lässt sich besser als critical reflection bezeichnen, um die Konzepte zu trennen (siehe Abschnitt 2.7.2). Die dritte Art bezieht sich auf die reflection-in-action von Schön (siehe Abschnitt 2.4.2; Schön 1983). Somit ist der Begriff der reflexivity bei Freshwater und Rolfe in der Definition zu breit, um in der Praxis nützlich zu sein (Freshwater und Rolfe 2001).

Es lässt sich bei anderen Autoren herausarbeiten, dass Reflexion sich eher auf Nachdenken über ein Objekt bezieht und reflexivity einen eher unmittelbaren Charakter hat und sich auf einen selbst bezieht (Finlay 2002; Swan 2008; West 1996). Diese Elemente finden sich bspw. bei Boud et al. (siehe Abschnitt 2.4.4; Boud, Keogh, and Walker 1985).

Auf der sozialen Ebene wird von *team reflexivity* gesprochen, welches sich auf eine Gruppe bezieht, die ein Fokus auf die eigenen Gruppenprozesse und ihre interne und externe Umgebung hat (West 1996; Hoegl und Parboteeah 2006). West sagt, dass diese Form nicht sehr weit verbreitet ist (West 1996). Für Hoegl und Parboteeah ist die

Fähigkeit als Team reflexivity zu zeigen wichtig für den Erfolg des Teams und sie zeigen in einer Studie, dass team reflexivity einen positiven Einfluss auf die Effektivität und die Effizienz der Teams hat (Hoegl und Parboteeah 2006). Hoegl und Parboteeah nutzen den Begriff als Eigenschaft eines Teams, welches konstant reflektiert. Der Unterschied zu der zuvor beschriebenen kollaborativen Reflexion (siehe Abschnitt 2.5) liegt also darin, dass bei team reflexivity der Fokus auf den Gruppenprozessen und der Umgebung selbst liegt und bei kollaborativer Reflexion der Fokus auf Problemen Einzelner liegt, die sie gerne in einer Gruppe reflektieren möchten. Der ablaufende Prozess scheint der Gleiche zu sein und der Unterschied liegt eher im Schwerpunkt dessen, worüber reflektiert wird.

Es lässt sich also feststellen, dass reflexivity bzw. team reflexivity eher auf die Reflexion über einen selbst bezieht und nicht auf die Reflexion über bspw. eine Tätigkeit oder Handlung.

### **2.7.5 Sensemaking**

Der Begriff des *sensemaking* wird bei Bell als das Verstehen der eigenen Argumente zu einem Thema durch das Auseinandersetzen mit dem Thema verstanden (Bell 2000). Bei Weick wird sensemaking als das Verstehen von unerwarteten Ereignissen verstanden, bei dem jemand zu dem Ereignis zurückblickt, dieses versucht zu verstehen und Theorien entwickelt, warum der Sachverhalt so existiert (Weick 1995, 1–16; oder Kellas und Trees 2006). Mit dieser Beschreibung liegt sensemaking damit sehr dicht bei dem Begriff der Reflexion. Prilla et al. sehen den Unterschied zur Reflexion darin, dass Reflexion einen klaren Fokus auf die Planung künftiger Schritte und Verhaltensweisen hat, welcher bei sensemaking oft nicht gegeben bzw. schwach ausgeprägt ist (Prilla, Herrmann und Degeling 2013; Prilla u. a. 2015). Diese Distanzierung passt zur Verwendung des Begriffs sensemaking bei anderen Autoren (Puusaar, Clear und Wright 2017; Ploderer u. a. 2014). Für Weick ist sensemaking ein sozialer Prozess (1995; Kellas und Trees 2006). Zusätzlich spricht Weick bei sensemaking davon, dass dies einen Fokus auf den Prozess des Verstehens selbst hat, was er von Interpretationen abgrenzt (1995). Dieser Fokus ist bei Reflexion nicht sehr ausgeprägt.

## **2.8 Unterstützungsbedarf und Unterstützungsmöglichkeiten von (kollaborativer) Reflexion**

Obwohl der Prozess der Reflexion sehr einfach aussehen mag, ist es dennoch ein Prozess, den viele Lernende herausfordernd finden und für den sie in der Praxis Unterstützung benötigen (Walsh 2009). In diesem Abschnitt werden unterschiedliche Faktoren

aufgeführt, die Reflexion in der täglichen Praxis erschweren und Möglichkeiten aufgezeigt, wie man Reflexion in der Praxis unterstützen kann.

### **2.8.1 Herausforderungen in der täglichen Praxis**

Ein häufiger Faktor, der Reflexion in der täglichen Praxis erschwert, ist die fehlende Zeit (bspw. Renner u. a. 2016; Hatton und Smith 1995; Moon 1999, 166; Fleck und Fitzpatrick 2010). Aus diesem Grund empfehlen Boud et al. die Reflexion in Lehr-Lernsituationen direkt in Übungen einzubauen, damit Zeit für Reflexion reserviert wird (Boud, Keogh und Walker 1985, 26). In einer Studie, bei der es um die Ausbildung von Physiotherapeuten ging, hat ein Ausbilder begonnen die Autofahrten zwischen zwei Patienten mit seinem Studenten für Reflexion zu nutzen (Morris und Stew 2007). Dies erlaubt es ihm in einem geringen Zeitabstand zu der eigentlichen Erfahrung, dem Umgang mit dem Patienten, zu reflektieren, ohne dass dadurch gesondert Zeit benötigt wird, die eventuell bei der Pflege anderer Patienten fehlen könnte. Hatton und Smith bestätigen, dass für die Reflexion nicht nur Zeit, sondern auch Gelegenheiten benötigt werden (Hatton und Smith 1995).

In eigenen Studien wurde beobachtet, dass Lernen in der Organisation geschätzt wurde, aber nicht im Konflikt zum Kerngeschäft des Unternehmens stehen durfte, sprich nicht zu viel Zeit in Anspruch nehmen durfte (Blunk und Prilla 2017a). Prilla et al. haben beobachtet, dass Krankenschwestern häufig zwischen Aufgaben oder während einer Übergabe zwischen zwei Schichten alleine oder kollaborativ reflektieren (Prilla, Degeling und Herrmann 2012). Es wurde ebenfalls beobachtet, dass viele Personen nach Ende ihrer Schicht beginnen zu reflektieren. Aus dem Grund fassen Prilla et al. zusammen, dass Unterstützung für Reflexion dementsprechend flexibel sein muss. Mangelnde Zeit für Reflexion könnte möglicherweise auch auf ein Problem der Priorisierung hindeuten, was beinhaltet, dass innerhalb der Organisation die Priorität für die Reflexion der eigenen Tätigkeit nicht besonders hoch ist bzw. andere Aufgaben aktiv in der Dringlichkeit höher bewertet werden.

Ein weiteres häufiges Problem ist das fehlende Festhalten von Ergebnissen aus häufig spontanen Reflexionen, da in der Praxis oft die nächste Aufgabe wartet (Prilla, Degeling und Herrmann 2012). Die Ergebnisse können bspw. über Meetings verbreitet werden (Prilla, Herrmann und Degeling 2013).

Moon fasst unterschiedliche Eigenschaften einer guten Reflexionsumgebung zusammen (vgl. für diesen Abschnitt Moon 1999, 165–176). Neben den beschriebenen Faktoren Zeit und Geld kommt bei Moon u. a. noch die Unterstützung der Gemeinschaft (im Orig.

*emotionally supportive environment*) dazu, womit sie eine Umgebung meint, in der sich Einzelne wohl genug fühlen sich zu öffnen und ihre Gedanken zu einem Sachverhalt entwickeln können. Moon hebt hervor, dass Hierarchien für Reflexion negativ sein können (siehe Abschnitt 2.5.4). Sie beschreibt ebenfalls, dass es hinderlich wird, wenn es Rezepte für Reflexion gibt, die jedes Mal gleich lauten und dann eher uninspiriert ausgefüllt werden, anstatt echtes Lernen anzuregen. In Gruppen kann das Problem auftauchen, dass sich manche versuchen zu verstecken und so (un)bewusst ihre eigene Reflexion behindern.

In der Praxis wird zudem nicht oft sehr systematisch reflektiert (Cheetham und Chivers 2000; Renner u. a. 2019), was die zuvor beschriebenen Ebenen von allgemeiner und akademischer Reflexion bestätigt (siehe Abschnitt 2.4.6). Zusätzlich benötigt Reflexion in der Praxis Unterstützung hinsichtlich der Strukturierung (Fleck und Fitzpatrick 2009; Renner u. a. 2019). Weiterhin existiert das Paradox, dass Lernsituationen erschwert werden, wenn sie explizit als Lernsituation beschrieben werden und die Empfehlung ist diese als Sitzungen für Projektentscheidungen, Evaluationen, oder Verbesserungen zu benennen (Boud, Cressey und Docherty 2006, 202). Zusätzlich benennen Boud et al. das Paradox, dass es vielen Personen schwerfällt auf der einen Seite kompetent zu wirken und auf der anderen Seite Gelegenheiten zum Lernen zu benötigen (Boud, Cressey und Docherty 2006, 202). Der Bedarf, Gelegenheiten zu schaffen, wird bei Dyke bestätigt (Dyke 2006).

Um Möglichkeiten und Freiräume zur Reflexion in der Praxis zu schaffen, wird Unterstützung seitens des Managements benötigt (Clouder 2000; Moon 1999, 165–176). Gleichzeitig darf aber Reflexion nicht eingefordert werden, da es sonst auf Ablehnung stößt (Boud, Cressey und Docherty 2006, 202).

Um Reflexion in der Praxis zu ermöglichen, sollte man häufiger aktiv Barrieren entfernen, als Reflexion nur zu ermutigen (Moon 1999, 174). Wobei es Autoren gibt, die bemerken, dass trotz Unterstützung von Reflexion dennoch die Motivation ein Problem ist (Clarà 2015). Selbst wenn das Wissen vorhanden ist, wie Reflexion anzuwenden bzw. durchzuführen ist, impliziert dies nicht, dass man es in der Praxis anwenden kann (Platzer, Snelling und Blake 1997). Ähnlich ist es bei Systemen, die Daten über persönliche Aktivitäten aufzeichnen (siehe Abschnitt 2.3), da zwar oft gezeigt wird, wie das aktuelle Verhalten aussieht, aber eine Unterstützung fehlt, darüber gezielt zu reflektieren und sein Verhalten in der Zukunft zu ändern (Saksono und Parker 2017).

Zusammenfassend kann festgestellt werden, dass es in der Praxis unterschiedlichste Faktoren gibt, die das Lernen am Arbeitsplatz erschweren. Dazu gehören bspw.

mangelnde Zeit (bspw. Renner u. a. 2016; Hatton und Smith 1995; Moon 1999, 166; Fleck und Fitzpatrick 2010), mangelnde Gelegenheiten (Hatton und Smith 1995), Ablenkung durch andere dringende Aufgaben (Prilla, Degeling und Herrmann 2012), fehlende Strukturierung (Fleck und Fitzpatrick 2009; Renner u. a. 2019), oder fehlende Motivation (Clarà 2015). Hieraus lässt sich ableiten, dass kollaborative Reflexion in der Praxis unterstützt werden muss, damit sie in der Praxis auch durchgeführt wird und dafür angemessene Freiräume in Form von Gelegenheiten (bspw. Zeit) und Möglichkeiten (bspw. Ort) geschaffen wird. Unterschiedliche Formen möglicher Unterstützung (kollaborativer) Reflexion werden in den nächsten Abschnitten beschrieben.

### **2.8.2 Unterstützungsmöglichkeiten (kollaborativer) Reflexion**

Nach dem der allgemeine Bedarf an Unterstützung für (kollaborative) Reflexion festgestellt wurde, wird in diesem Abschnitt die Passung computergestützter Unterstützungsmöglichkeiten beschrieben.

Um Reflexion in der Praxis zu unterstützen und um Reflektierenden mehr Nutzen aus ihrer Reflexion zu ermöglichen, wird oft computergestützte Reflexion propagiert (Prilla, Herrmann und Degeling 2013; Cheetham und Chivers 2000; Høyrup 2004; Renner u. a. 2019). Es wird allerdings reflexionsspezifische Unterstützung benötigt und es genügt nicht nur, eine Software zum Austausch von Informationen bereitzustellen (Kimmerle u. a. 2011; Renner u. a. 2016). Oft schlägt die Software in Studien allerdings den Nutzern keine konkreten Schritte für eine Umsetzung vor (Baumer u. a. 2014). Renner hebt ebenfalls hervor, dass dies sowohl auf organisationale Kontexte (Kimmerle und Cress 2009) als auch Online-Foren (Oliveira, Tinoca und Pereira 2011) zutrifft. Es wird ebenfalls personalisierbare Unterstützung benötigt, da Lerner unterschiedlich in ihren Bedürfnissen bzw. in ihren Lernansätzen sind (Kolb 1984; siehe Abschnitt 2.4.3).

Wichtige zu unterstützende Aspekte bei (kollaborativer) Reflexion, sind das Erfassen und Teilen eigener Erfahrungen, damit man über diese alleine oder gemeinsam reflektieren kann (Prilla, Herrmann und Degeling 2013; Fleck und Fitzpatrick 2010).

Zur Unterstützung von Reflexion werden unterschiedlichste Mechanismen eingesetzt, die im Folgenden kurz erläutert werden. Der Autor hat an einer umfassenden systematischen Analyse aktueller Unterstützungsmechanismen mitgewirkt (Fessler u. a. 2017). In der Analyse werden die folgenden Mechanismen bspw. nach der zeitlichen Dimension untersucht, ob sie sich zur Unterstützung von reflection-in-action und/oder reflection-on-action eignen; ob sich die Mechanismen zur Unterstützung von individueller oder kollaborativer Reflexion eignen; und welche Arten von Führung (im Orig. *Guidance*) die



Mechanismen für Reflexion bieten. Alle Unterstützungsmechanismen in den folgenden Abschnitten sind computergestützt.

### 2.8.3 Journale

Schreiben ist ein beliebtes Mittel, um zu reflektieren, und wurde früh beschrieben (Walker 1985). Walker spricht in dem Kapitel über Portfolios, beschreibt aber eine Art papierbasiertes Journal, was zu der folgenden Beschreibung von Moon passt. Moon weicht in ihrem Buch einer konkreten Definition von Journalen für Reflexion aus, und schreibt stattdessen von *learning journal*, womit sie eine Sammlung von Material meint, dass auf dem Reflexionsprozess des Lernalters basiert<sup>6</sup> (Moon 2006b, 2). Ein Journal kann laut Moon, sowohl alleine als auch gemeinsam geschrieben werden (Lee 2005). Letzteres wird bei ihr als *Dialogue Journals* beschrieben und bedeutet, dass jeder zu den Einträgen der anderen Teilnehmer Stellung nimmt. Journale werden bspw. häufig im Bereich der Krankenpflege eingesetzt (Platzer, Snelling und Blake 1997; Sandars 2009; Chirema 2007).

Oft werden Journale zur Reflexion in Trainingssettings eingesetzt. Es ist beispielsweise nicht klar, ob diese bewertet werden sollen (Chirema 2007). In einer Studie zeigen Sinclair und Tse, dass sich Journals sehr gut eignen, um Studenten ein Werkzeug zur Reflexion an die Hand zu geben, dass nach Meinung der Autoren die Studenten in der Studie dazu befähigt hat, um später in der Praxis weiter zu reflektieren (Sinclair und Tse 2008). Shields bestätigt, dass sich reflexives Schreiben sehr gut dazu eignet Studenten Reflexion zu vermitteln und spricht davon, dass die Studenten dies sehr geschätzt haben (Shields 1995). Dies hat sich ebenfalls in den Studien von Loo und Thorpe gezeigt (2002). Hatton und Smith geben zu bedenken, dass in ihren Studien Studenten häufig auf einem deskriptiven Level der Reflexion stehen blieben und die wenigsten Studenten eine Ebene erreicht haben, auf der sie tiefergehend bzw. abstrakt über Probleme reflektiert haben (Hatton und Smith 1995). Wopereis et al. bestätigen das und unterstreichen, dass eine intensive Anleitung für reflexives Schreiben von Nöten ist, damit Journale gut funktionieren (Wopereis, Sloep und Poortman 2010).

Nachteilig an Journals sind sowohl der Aufwand und die Schwierigkeit die Einträge zu verfassen (Chirema 2007). Man kann die Schreibenden durch reflexive Fragen oder durch gegenseitige Unterstützung fördern (Loo und Thorpe 2002). Dennoch ist reflexives

---

<sup>6</sup> Im Orig.: „[...] we refer to an accumulation of material that is mainly based on the writer's process of reflection“

Schreiben eine Technik, die in vielen Studien zu reflexivem Lernen geführt hat (vgl. Übersicht in Fessler u. a. 2017; und bspw. Sinclair und Tse 2008; Chirema 2007; Shields 1995; Loo und Thorpe 2002). Nachteilig ist auch, dass manche Studenten teils höheren Betreuungsaufwand benötigen (Chirema 2007). Dadurch ist dieses Hilfsmittel in größeren Gruppen schwierig zu skalieren.

Abgrenzen lassen sich Journale unter anderem von Tagebüchern, die mehr zum persönlichen Lernen eingesetzt werden (Tang 2000), und nicht mit anderen geteilt werden (Fessler u. a. 2017). Dennoch können Journale für persönliches Lernen eingesetzt werden (Moon 2006b, 3). Ein Journal muss nicht immer ein schriftlicher Eintrag sein, da Moon zusätzlich Sprachaufnahmen als Möglichkeit zur Reflexion verwendet (Moon 2006b, 3). Reflexive Journale können in der Form von Blogs angelegt werden, was für Studenten und angehende Lehrer in der Ausbildung zur Unterstützung der Reflexion förderlich war (Yang 2009; Wopereis, Sloep und Poortman 2010; Stoszkowski und Collins 2017).

#### **2.8.4 E-Portfolios**

Als E-Portfolio wird eine Sammlung von Artefakten oder Informationen bezeichnet, die eine Person zu Lernzwecken über Zeit in einem Onlinetool ansammelt (Alexiou und Paraskeva 2010; Elango, Jutti und Lee 2005). Daher ist ein E-Portfolio eine Form reflexiven Schreibens, wie die Strukturierung in Fessler et al. nahelegt (Fessler u. a. 2017). Allerdings kann ein Portfolio zusätzlich Bilder (siehe Abschnitt 2.8.5), Grafiken oder Datenreihen aufnehmen und verwalten (Land und Zembal-Saul 2003). Am geeignetsten lässt sich ein E-Portfolio nutzen, wenn es bspw. in eine Trainingseinheit eingebettet und dort regelmäßig genutzt wird (Alexiou und Paraskeva 2010).

Bei Elango et al. nimmt das E-Portfolio beispielsweise die Form an, dass Studenten über einen Zeitraum von mehreren Semestern schriftliche Berichte abgeben mussten (Elango, Jutti und Lee 2005). Das E-Portfolio beinhaltet dann alle Berichte und hilft beim weiteren Lernen. Über die Hälfte (65 %) der Teilnehmer in der Studie haben berichtet, dass sie E-Portfolios als Werkzeug nützlich finden. Man kann während oder erst am Ende des Kurses über die Inhalte des E-Portfolios reflektieren (Lee 2005; vgl. im Folgenden Scott 2010). In einer Studie hat Scott gezeigt, dass die Studenten, die regelmäßig ihr E-Portfolio verwendet haben, am Ende des Kurses höhere Werte für Reflexion hatten. Scott hebt hervor, dass es wichtig ist, in formalen Trainings das E-Portfolio einzubeziehen, da der Sinn zweifelhaft ist, wenn die Studenten mit dem Tool nicht üben können und es nur am Ende des Kurses verwenden.

Eine Herausforderung im Einsatz von E-Portfolios ist das Kaltstart-Problem, wonach Studenten beim Aufbau ihres E-Portfolios aufgrund dessen, dass sie das zu lernende Thema noch nicht gut kennen, kein Bild davon haben, welche Dokumente oder Informationen wertvoll genug sind, um sie in das E-Portfolio aufzunehmen (Land und Zembal-Saul 2003). Als weiteren Nachteil nennt Scott, das E-Portfolios ein sehr zeitaufwendiges Mittel der Pädagogik sind, da es von Seiten der Dozenten betreut werden muss und die Studenten viel Zeit aufwenden, um Inhalte in ihr Portfolio zu bringen bzw. dort zu erstellen (Scott 2010).

### **2.8.5 Visualisierungen**

Eine Möglichkeit, um Reflexion zu unterstützen, sind Visualisierungen eigener Daten, die im Rahmen von Quantified Self angefallen sind (siehe Abschnitt 2.3). Dies hilft beispielsweise bei der Analyse von Daten über physische Aktivitäten, wie Schritte zählen oder die Analyse des eigenen Schlafs (Choe u. a. 2017). Gleichzeitig helfen diese Visualisierungen im medizinischen Bereich, bspw. bei der Visualisierung von Blutzuckerwerten (Mamykina u. a. 2016). Im Lehr-Lernkontext helfen Bilder eigene Tutorien zu analysieren und zu reflektieren, und bieten eine gute Möglichkeit den Kontext der Reflexion durch das Bildmaterial zu schärfen (Fleck und Fitzpatrick 2009; Fleck und Fitzpatrick 2006). Nachteilig an den Studien von Fleck und Fitzpatrick ist, dass die Fotoaufnahmen kontinuierlich und nicht bei Bedarf angefertigt werden und dementsprechend andere Personen im Sichtbereich des Nutzers aufgenommen werden. In diesem Falle ist zusätzlich die Berücksichtigung der Privatsphäre Dritter wichtig.

In einer Studie haben Leijen et al. Videoaufzeichnungen von Ballettstudenten genutzt, um diese über ihre Trainingseinheiten reflektieren zu lassen (vgl. im Folgenden Leijen u. a. 2009). Die Studenten haben die externe Sichtweise auf die eigenen Handlungen geschätzt und positiver als die Korrekturen des Trainers gewertet.

Neben Bildern und Daten über eigene physische Aktivitäten bzw. Gesundheitsdaten, nennt Prilla zusätzlich Sensordaten (bspw. gemessene Stimmung), Workflowdaten, Anwendungsdaten (geteilte Bibliotheken bzw. Lesezeichen), Notizen und Dokumentationen der Tätigkeit als nützliche Artefakte zur Reflexion (Prilla u. a. 2011; Lin u. a. 1999). Lin et al. schlagen Anzeigen für Lernprozesse vor, auf denen man sehen kann, welche Tätigkeiten man aktuell ausführt, um ein Problem zu lösen (Lin u. a. 1999). Laut Lin et al. hilft dies, vom Prozess etwas Abstand zu nehmen und über den Prozess zu reflektieren.

Nachteilig bei der Reflexion, basierend auf Artefakten, wie Bildern oder Videos, ist, dass der Reflektierende die passenden Artefakte erst in der Masse auffinden muss. Auf der anderen Seite bietet das Stöbern in den Bildern und Videos eine Möglichkeit zur Reflexion, da man an Situationen erinnert wird, die man ggf. bereits vergessen hat.

Visualisierungen können ein Element sein, welches in E-Portfolios gespeichert wird (siehe Abschnitt 2.8.4).

### **2.8.6 Prompts**

Bei Prompts handelt es sich nicht um ein spezifisches Tool, wie beispielsweise E-Portfolios (siehe Abschnitt 2.8.4), sondern mehr um ein Hilfsmittel, um Reflexion anzuregen (Fessler u. a. 2017). Prompts sind oft textuelle Hinweise (Bannert 2009), um Erinnerungen auszulösen (Chen u. a. 2009) oder Aspekte von Reflexion in den Vordergrund zu rücken.

Der Hauptvorteil von Prompts ist, dass sie rein maschinell selektiert und ausgeliefert werden und nicht, wie die vorherigen Hilfsmittel, menschliche Betreuung benötigen. Dadurch sind Prompts ein Mechanismus zur Reflexionsunterstützung, der sehr gut skaliert. Prompts werden im Abschnitt 2.10 ausführlich beschrieben.

## **2.9 Gestaltung von Reflexionsunterstützung**

Nachdem im vorherigen Abschnitt (siehe Abschnitt 2.8) festgestellt wurde, dass für Reflexion eine Unterstützung in der Praxis notwendig ist, fehlt noch eine Annäherung daran, wie diese Unterstützung aussehen kann.

### **2.9.1 Modell zur Unterstützung des Reflexionsprozesses**

In der vorliegenden Arbeit soll kollaborative Reflexion durch Software unterstützt werden. Viele Modelle, wie bspw. das zuvor erwähnte Modell (siehe Abschnitt 2.5.2) kollaborativer Reflexion von Prilla et al. (2012), beschreiben Reflexion, aber lassen zunächst eine mögliche Unterstützung durch Technologie außen vor. Dazu passend gibt es ein Modell von Krogstie et al., siehe Abbildung 2-4, welches phasenorientiert ein Modell von kollaborativer Reflexion beschreibt (vgl. im Folgenden Krogstie, Prilla und Pammer 2013):

In der ersten Phase („*Plan and do work*“, siehe Abbildung 2-4) wird die reguläre Arbeit am Arbeitsplatz getätigt, geplant und beobachtet (sei es durch einen selbst oder durch

automatisierte Tools). In der zweiten Phase („*Initiate reflection*“) wird die Reflexion an sich angestoßen, indem jemand das Problem anspricht und das Ziel für die Reflexion nennt. Bei Bedarf werden andere Personen hinzugezogen, was in dem Modell als Schnittstelle zwischen individueller und kollaborativer Reflexion dient. Es wird an dieser Stelle die eigentliche Reflexionssitzung geplant. Die dritte Phase („*Conduct reflection session*“) beinhaltet die eigentliche Reflexionssitzung. Es können eine oder mehrere Personen über das Thema reflektieren, gegenseitig Erfahrungen austauschen und an einem gemeinsamen Verständnis der Situation arbeiten. Zusätzlich erarbeitet man in dieser Phase ein Ergebnis (im Orig. *outcome*) und plant, wie dieses in der Praxis einsetzbar ist. Das Überführen des Ergebnisses in die Praxis stellt die vierte Phase („*Apply Outcome*“) dar. Es wird allein oder mit anderen versucht das Ergebnis zu implementieren. Das Ziel ist es, das Ergebnis tatsächlich in die Praxis zu bringen. Bei Erfolg oder bei Misserfolg werden neue Erfahrungen gesammelt, die wiederum dazu genutzt werden können, den Zyklus von vorne zu beginnen.

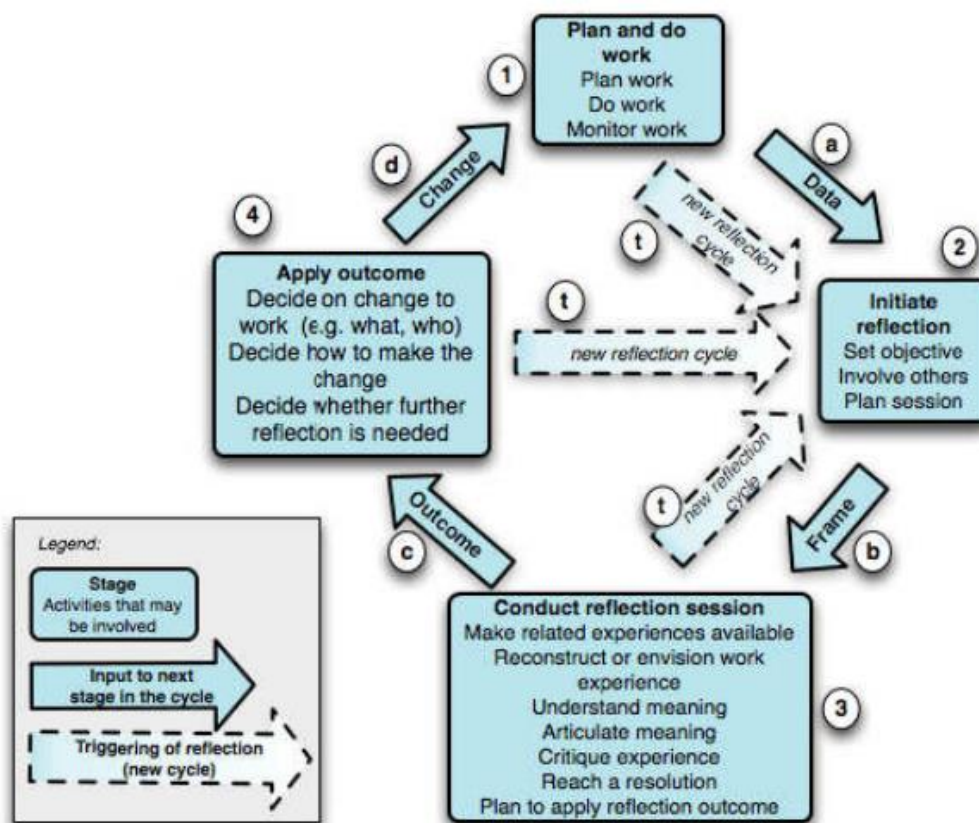


Abbildung 2-4: Computer-supported reflective learning model nach Krogstie et al. (2013, Fig. 1)

Um besser einen Prozess unterstützen zu können, besitzt das CSRL-Modell, siehe Abbildung 2-4, Übergänge zwischen den einzelnen Phasen (vgl. im Folgenden Krogstie, Prilla und Pammer 2013). Zwischen den ersten beiden Phasen werden Daten genutzt, die aus der Arbeit entstanden sind und ein Auslöser bzw. Gegenstand für die Reflexion sein

können. Nach der zweiten Phase existiert ein Rahmen (im Orig. *Frame*) für die Reflexion, welcher unterschiedlich detailliert sein kann, falls bspw. jemand nur plant jemanden anderen zu einem Problem anzusprechen. Das Ergebnis der dritten Phase bzw. der Reflexion, stellt den Input für die vierte Phase dar und beinhaltet den konkreten Plan dessen, was man in der Praxis demnächst umsetzen möchte. Nachdem in der vierten Phase das Ergebnis angewendet wurde, entsteht ein Wandel in der Praxis (im Orig. *Change*). Dieser Wandel betrifft die geänderten Tätigkeiten während der Arbeit.

Es ist ebenfalls festzustellen, dass in jeder Phase ein Rücksprung auf das Initiieren einer neuen Reflexionssitzung (Phase 2) ist, welches im Paper als *trigger* bezeichnet wird (vgl. im Folgenden Krogstie, Prilla und Pammer 2013). Dies kann bspw. abbilden, falls jemand allein reflektiert und merkt, dass er zu keiner Lösung kommt (Phase 3) und sich daher entscheidet jemanden anderen einzubeziehen (Phase 2), um dann gemeinsam zu reflektieren (Phase 3). Dies tritt aber nicht nur auf, wenn andere zur Problemlösung einbezogen werden, sondern auch, wenn andere einbezogen werden, um die Lösung des Problems zu verbreiten und dann weiter zu diskutieren.

Die Zielsetzung des Modells ist die Unterstützung bei der Gestaltung von computergestützter Reflexion. Es bietet dafür ein Phasenmodell, an dem sich die Entwickler orientieren können, um Reflexionsprozesse in ihrer Software zu unterstützen. Die Herausforderung liegt darin, zu erkennen, in welcher Phase eine Einzelperson oder eine Gruppe im Reflexionsprozess aktuell ist.

### **2.9.2 Gestaltung**

In diesem Abschnitt werden unterschiedliche Maßnahmen dargestellt, mit denen Reflexionsprozesse für Einzelpersonen oder Gruppen in der Praxis digital unterstützt werden können. Die häufigste Rolle der Technologie ist die Erfassung von Daten und Erfahrungen, welche dann die Position einnimmt Reflexion unterstützen zu können (im Orig. *enable*) (Prilla, Degeling und Herrmann 2012; Fleck und Fitzpatrick 2010).

Da man über eigene Erfahrungen reflektiert (siehe Abschnitte 2.4 und 2.5), wird zunächst eine Möglichkeit benötigt, die eigenen Erfahrungen zu erfassen und zu betrachten (Baumer 2015; Fleck und Fitzpatrick 2010). Dadurch wird gleichzeitig das Erinnerungsvermögen des Reflektierenden unterstützt (Fleck und Fitzpatrick 2010). An dieser Stelle passen unter anderem Journale (siehe Abschnitt 2.8.3) und E-Portfolios als Tool für so eine Unterstützung, da man dort Daten unterschiedlichster Art (bspw. Notizen, Bilder, etc.) speichern kann (siehe Abschnitt 2.8.4). Das Erfassen von Daten und Erfahrungen in einem Tool kann helfen, Reflexion von Angesicht zu Angesicht zu fördern

(Prilla und Renner 2014). Aufgrund der fortschreitenden Digitalisierung im beruflichen und privaten Bereich kann die Reflexion zu den Erfahrungen hinter den Daten geschehen, die technologisch erfasst werden (Fleck und Fitzpatrick 2010). Technologie kann Daten erfassen, die der Mensch selbst schwer erfassen kann (Fleck und Fitzpatrick 2010; Slovák, Frauenberger und Fitzpatrick 2017; Marcu, Dey und Kiesler 2014), wie bspw. Pulsmessgeräte bei physischen Aktivitäten, wie Jogging.

Zusätzlich kann Technologie eine Möglichkeit bereitstellen, das eigene Material umzuorganisieren und aus unterschiedlichen Blickwinkeln zu betrachten, da dies zu Reflexion führen kann (Fleck und Fitzpatrick 2010; Mols, van den Hoven und Eggen 2016a). Baumer empfiehlt eine Trennung der Erfahrungen und einer Möglichkeit sich mit den Erfahrungen reflexiv auseinanderzusetzen (Baumer 2015). Slovák et al. beschreiben einen ähnlichen zweiseitigen Prozess (Slovák, Frauenberger und Fitzpatrick 2017).

Fleck und Fitzpatrick heben hervor, dass die Technologie der Zielsetzung der Reflexion folgen soll und zunächst zu klären ist, welches reflexive Verhalten der Nutzer an den Tag legen soll (Fleck und Fitzpatrick 2010; oder Lin u. a. 1999; Lee 2005). Ein weiterer Vorschlag ist, dass nicht nur die Reflektierenden unterstützt werden, sondern auch ggf. die Mentoren oder Dozenten, die die Reflektierenden betreuen bzw. unterstützen (Slovák, Frauenberger und Fitzpatrick 2017; Lin u. a. 1999). Motivation ist ein Faktor, der bspw. über Gamification gefördert werden kann (Slovák, Frauenberger und Fitzpatrick 2017).

Beim Auslösen von Reflexion ist es wichtig, darauf zu achten, dass der Nutzer nicht bei wichtigen Tätigkeiten gestört wird (Mols, van den Hoven und Eggen 2016a). Laut Mols et al. gibt es am Tag häufiger Situationen, in denen keine Zeit oder Gelegenheit für Reflexion möglich ist, was bedeutet, dass Technologien eng an den Tagesablauf bzw. Alltag angelehnt sein müssen. Es bietet sich laut Mols et al. daher an, die Reflexion mit bestimmten Ritualen zu verbinden, wie bspw. die MirrorMirror-Installation, die mit täglichen Zeiten im Bad koppelbar ist.

Die Möglichkeit, eigene Reflexionen zu einem Thema mit anderen zu teilen bzw. in eine kollaborative Reflexion überzugehen (Baumer 2015; Slovák, Frauenberger und Fitzpatrick 2017; Mols, van den Hoven und Eggen 2016a; Prilla und Renner 2014; Lin u. a. 1999; Lee 2005), wodurch dann in einer Gruppe reflektiert werden kann. Bei geteilten Daten werden Gelegenheiten geschaffen, dass andere spontan über die geteilten Daten reflektieren können (Lin u. a. 1999). Prilla weist darauf hin, dass die Gruppengröße einen Einfluss auf Reflexion haben kann, da bei kleinen Gruppen häufig nicht jeder verfügbar ist oder manche keine Kenntnis über die konkrete Problemsituation haben (Prilla 2014; Prilla 2015). Eine Empfehlung ist die Implementierung einer Möglichkeit in

größeren Gruppen, um bspw. mit Externen in Communities (siehe Abschnitt 2.1) zu reflektieren (siehe auch Lin u. a. 1999). Wichtig für das Reflektieren in einer Gruppe, zu der Vorgesetzte gehören (siehe Abschnitt 2.5.4), scheint die Möglichkeit, anonyme Beiträge in einer Diskussion schreiben zu dürfen (Prilla und Hermann 2017). Nicht nur das Teilen von erfassten Erfahrungen ist interessant, sondern auch die Möglichkeit eine Diskussion darüber zu führen, wie man am besten Erfahrungen erfassen kann (Marcu, Dey und Kiesler 2014).

Ein Unterstützungsmittel für Reflexion können Fragen sein, die den Reflektierenden zum Nachdenken anregen (Daudelin 1996; Fleck und Fitzpatrick 2010; Slovák, Frauenberger und Fitzpatrick 2017). Prompts sind ein Mittel diese Fragen zu stellen und werden im folgenden Abschnitt über Prompts ausführlicher beschrieben (siehe Abschnitt 2.10).

Obwohl Reflexion allgemein positive Effekte zugeschrieben werden, gibt es teils negative Effekte, falls jemand nicht konstruktiv über einen längeren Zeitraum bei negativen Erfahrungen verharrt und daher empfehlen Mols et al. unter anderem Funktionen einzubauen, mit der Technologien Reflexion bei Bedarf unterbrechen oder konstruktiver gestalten können (Mols, van den Hoven und Eggen 2016b).

Häufig ist es ein Problem, neue Software in einem Unternehmen einzuführen, mit denen Personen reflektieren können, da Mitarbeiter diese Tools erst verwenden, wenn ein Nutzen demonstriert wurde (Prilla und Hermann 2017). Ein Ansatz kann sein, Reflexionsunterstützung über Plug-ins zu existierenden Tools hinzuzufügen (vgl. im Folgenden Prilla und Blunk 2015). Auf diese Weise kennen die Nutzer die Tools bereits und es ist vermutlich leichter, aus dem Arbeitskontext in den bisherigen Tools heraus einen neuen Reflexionsprozess zu starten. Es können unterschiedliche Mechanismen in die existierenden Anwendungen gebracht werden, wie bspw. Prompts (siehe Abschnitt 2.10) oder Vorschläge, welche anderen Kollegen bzw. Nutzer in Communities (siehe Abschnitt 2.1) einem in der Reflexion helfen können.

Die Technologie muss zum Arbeitskontext des Nutzers passen. Wenn ein Beruf überwiegend auf Kommunikation von Angesicht zu Angesicht beruht, dann ist es schwierig, ein Tool für digitale kollaborative Reflexion zu etablieren (vgl. für diesen Abschnitt Prilla und Hermann 2017). Je nach Beruf können unaufschiebbare Unterbrechungen, bspw. Notfälle im Krankenhaus, den Reflexionsprozess unterbrechen. Daher ist es wichtig, dass das Tool sowohl Unterbrechungen als auch Wiederaufnahme des Reflexionsprozesses unterstützt.

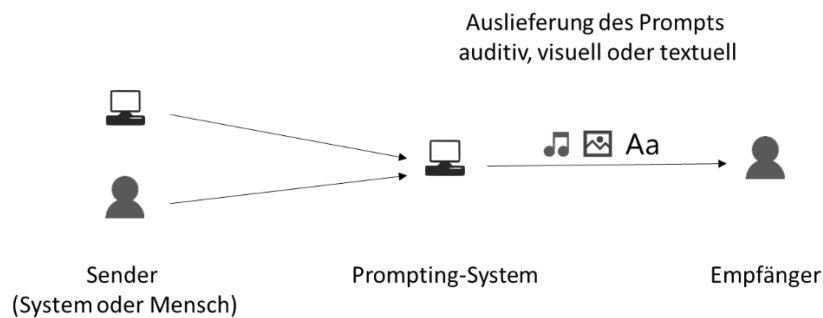


## **2.10 Prompts als skalierbare und personalisierbare Unterstützung für Reflexion**

In den vorherigen Abschnitten wurden Prompts bereits kurz als skalierbare Möglichkeit der (kollaborativen) Reflexionsunterstützung vorgestellt (siehe Abschnitt 2.8.6). In diesem Abschnitt wird das Konzept der Prompts näher beschrieben und von anderen Konzepten abgegrenzt, sowie der aktuelle Stand der Literatur aufgearbeitet.

### **2.10.1 Definition**

An dieser Stelle wird zunächst eine allgemeine Definition von Prompts angestrebt. In Abschnitt 2.10.4 werden unterschiedliche Arten von Prompts dargestellt. Ein Prompt stellt eine Hilfestellung dar, die ein Nutzer für einen spezifischen Kontext, in dem er aktuell ist, bekommt (Davis und Linn 2000; Lai und Calandra 2007; Davis 2003). Ein Prompt kann sowohl eine Frage (Fivush, Gray und Fromhoff 1987; Bannert 2009; Berthold, Nückles und Renkl 2007; van den Boom, Paas und van Merriënboer 2007; Saksono und Parker 2017; Clarke 2004; Papadopoulos u. a. 2009) als auch eine Aufforderung (Thillmann u. a. 2009; Bannert 2007), oder visuell oder auditiv sein (McKenzie-Mohr 2000; Chung und Byrne 2008). Ein Prompt kann ein Satzanfang sein, den der Empfänger vervollständigen muss, was bspw. im Bildungsbereich genutzt wird (Davis 2003; Clarke 2004). Die Hilfestellung kann dazu da sein, an eine Aktivität zu erinnern, die man andernfalls vergessen hätte (McKenzie-Mohr 2000; Bannert 2009). Bannert konkretisiert dies und sagt, dass Prompts eine Hilfe für Erinnerungen oder Unterstützung für die Durchführung einer Aktivität (im Orig. „*performance aids*“) sind und immer auf der Annahme basieren, dass der Empfänger weiß, was zu tun ist, aber sich in der Situation entweder nicht (exakt) erinnert oder konkrete Hinweise für die Durchführung benötigt (Bannert 2009; Glogger u. a. 2009; Herrmann und Nierhoff 2017). In dem Sinne werden Prompts nicht dazu verwendet, dem Empfänger neue Themen beizubringen, sondern helfen Gelerntes in Erinnerung zu rufen (Deitz und Malone 1985; Glogger u. a. 2009). Nach Skinner erhöhen zusätzliche Stimuli die Wahrscheinlichkeit, dass existierendes Verhalten gezeigt wird (Skinner 1957, 52–55). Deitz und Malone sehen Prompts als anfängliche Unterstützung, um fehlerfreies Lernen zu ermöglichen, da Prompts genutzt werden können, um korrekte Antworten zu erzielen (Deitz und Malone 1985). Dementsprechend können Prompts neben Fragen (Griffith und Frieden 2000) zusätzlich konkrete Instruktionen für Aktivitäten beinhalten (Thillmann u. a. 2009; Fivush, Gray und Fromhoff 1987). Stichter fasst zusammen, dass Prompts darauf abzielen eine Antwort zu bekommen oder eine Handlung auszulösen (Stichter u. a. 2009).



**Abbildung 2-5: Unterscheidung zwischen dem Sender, der über einen Prompting-Mechanismus einen auditiven, visuellen oder textuellen Prompt an den Empfänger schickt**

An dieser Stelle wird unterschieden zwischen dem Prompt, dem Hinweis (textuell, visuell, auditiv), dem Sender des Prompts, dem Empfänger eines Prompts und des Prompt-Mechanismus, der genutzt wird (siehe Abbildung 2-5). Der Sender nutzt den Prompt-Mechanismus, um den Prompt an einen Empfänger zu senden, um einen Effekt auszulösen. In diesem Fall wirkt der Prompt wie eine Nachricht (Herrmann und Nierhoff 2017). Prompts können in Erinnerungen für einen selbst genutzt werden, wodurch in dem Fall der Sender und der Empfänger die gleiche Person sind. Der Effekt kann, wie oben beschrieben, eintreten, weil der Empfänger den Prompt nicht befolgen muss.

Es ist wichtig zu unterstreichen, dass Prompts als Hilfestellung gedacht sind und dem Empfänger eine Möglichkeit der Reaktion geben, ihn aber nicht dazu zwingen auf eine bestimmte Art oder überhaupt zu reagieren (Fivush, Gray und Fromhoff 1987; Saksono und Parker 2017; Herrmann und Nierhoff 2017). Somit kann man Prompts von Gesetzen oder Regeln abgrenzen, an die sich ein Nutzer halten muss bzw. bei denen eine Strafe impliziert ist, falls dieser das nicht tut. Dieser Aspekt ist für die Reflexion wichtig, da Cressey et al. feststellen, dass zu starke Strukturierung dem Prozess selbst schaden kann (Cressey, Boud und Docherty 2006). Aus diesem Grund kann der Sender eines Prompts nicht davon ausgehen, dass der Empfänger aufgrund des Prompts eine Handlung durchführt oder anpasst. Zudem lässt sich so feststellen, dass Prompts intentional sind, sprich der Sender sie auf eine bestimmte Art und Weise gestaltet hat, dass der Empfänger sie sieht. Ein Prompt muss nicht zwingend elektronisch sein, sondern kann in Papierform auch eine Unterstützung erzielen (Fiorella und Mayer 2012).

Aus der bisherigen Literatur wird für Prompts in der vorliegenden Arbeit die folgende Definition verwendet: Prompts sind für den Nutzer sichtbar positionierte textuelle Hinweise, welche die Wahrscheinlichkeit erhöhen sollen, dass der Empfänger dazu ermuntert wird eine optionale bekannte Handlung durchzuführen bzw. seine Handlung an den optionalen Hinweis anzupassen. Durch unterschiedliche Inhalte können Prompts sowohl einzelne Aspekte anregen als auch helfen in mehreren Schritten durch Prozesse

zu navigieren. Die Handlung kann Reflexion sein, aber Prompts können auch allgemeiner bzw. zu anderen Zwecken genutzt werden.

### **2.10.2 Einsatz und Wirkung von Prompts**

Prompts werden beispielsweise eingesetzt, um Personen zum Treppensteigen zu ermuntern (Webb und Eves 2007). Zusätzlich können Prompts zum Anregen kreativer Ideen beim Design von neuer Hardware genutzt werden (Yoo u. a. 2013). Eine beliebte Nutzung von Prompts ist die Vermeidung bzw. Reduktion von sogenannten *post-completion errors*, sprich Fehler, die jemand macht, nachdem das Ziel in einem Prozess bereits erreicht wurde und der Prozess abgeschlossen ist (Chung und Byrne 2004). Dazu gehört bspw. das Vergessen des Originals im Kopierer nach Anfertigen einer Kopie (Chung und Byrne 2004). Visuelle und auditive Prompts können helfen solche Art Fehler zu verringern (Chung und Byrne 2004). Eine Übersicht über mehrere Studien zeigt, dass Prompts mit Hinweis auf die Gesundheit, Personen dazu motivieren können eher die Treppe als die Rolltreppe oder einen Lift zu benutzen (Nocon u. a. 2010).

Ein häufiger Einsatzbereich von Prompts ist die Unterstützung von Lernstrategien im Bildungsbereich (Thillmann u. a. 2009; Bannert 2009), wo Prompts das Lernen fördern. Im Bildungsbereich lässt sich feststellen, dass Studenten häufig *copy-paste*-Strategien anwenden, um auf Reflexions-Prompts zu antworten oder die Prompts ignorieren (Furberg 2009). Allerdings helfen Prompts den Studenten das eigene Vorgehen bzw. den Ansatz zu hinterfragen (Furberg 2009). Wenig überraschend können so verschiedene Arten von Prompts (siehe Abschnitt 2.10.4) zu verschiedenen Lerneffekten führen (Rosenshine, Meister und Chapman 1996). Benachrichtigungen via E-Mail haben die Möglichkeit, zu höherer Wahrscheinlichkeit zu führen relevante Lernvideos anzuschauen und diese bis zu 15 % länger zu betrachten (van Oldenbeek u. a. 2019).

Zusätzlich werden Prompts zur Unterstützung von Reflexion eingesetzt und führen dazu, dass die Empfänger reflexiver agieren (Lai und Calandra 2007; Davis und Linn 2000; Davis 2003; Chen u. a. 2009; Ifenthaler 2012; van den Boom u. a. 2004; Furberg 2009; Lee 2005). Die beiden Bereiche Bildung und Reflexion überlappen sich größtenteils, da der überwiegende Teil der Prompts für Reflexion im Bildungsbereich genutzt wird. Es gibt Studien in denen Prompts für Interpretation bzw. Hilfe in einem Experiment erfolgreich getestet wurden und den Studenten beim Lernen geholfen haben (Reid, Zhang und Chen 2003). Prompts für Reflexion in Verbindung mit Unterstützung durch Tutoren, helfen persönliche Kompetenzen im selbstgesteuerten Lernen (im Orig. *self-regulated*

*learning*<sup>7</sup>) aufzubauen (van den Boom u. a. 2004; van den Boom, Paas und van Merriënboer 2007). Prompts, die nach Begründungen für das eigene Vorgehen fragen, helfen Studenten Wissen auf Problemlösung zu übertragen (Lin und Lehman 1999). In den Studien von van den Boom et al. gab es allerdings nur reflexive Dialoge, wenn ein Tutor als Reflexionspartner den Studenten geantwortet hat. Somit ist die Interaktion in der Umgebung sehr vom Moderator abhängig. Es gibt ebenfalls Reflexionsunterstützung durch Prompts am Arbeitsplatz (Blunk und Prilla 2017b; Renner u. a. 2016; Prilla 2014; Fessl u. a. 2015). Es ist festzustellen, dass sich die bisherige Literatur überwiegend auf die Unterstützung von Reflexion Einzelner fokussiert (Renner u. a. 2016) und Prompts, die allgemein Reflexion unterstützen, nicht zwingend kollaborative Reflexion fördern (van den Boom u. a. 2004; van den Boom, Paas und van Merriënboer 2007). Einzelne Studien nutzen Prompts im Bildungsbereich, um eine Dyade, die gemeinsam an einem Computer sitzt, zur Reflexion anzuregen (Furberg 2009). Ohne die Unterstützung von Prompts kann Reflexion auf einem deskriptiven Level bleiben (Roberts 2009) (siehe Abschnitt 2.4.6).

Beispiele von Prompts zur Unterstützung von Reflexion sind:

1. “What could the author have done differently?” (Renner u. a. 2016)
2. “Have you been in a similar situation? What did you do?” (Renner u. a. 2016)
3. “What would you personally suggest as a solution? Why?” (Blunk und Prilla 2015)
4. “What did surprise you in your daily work lately? Why?” (Blunk und Prilla 2015)

Diese vier Beispiele von Prompts verfolgen unterschiedliche Ziele bei der Unterstützung von Reflexion. Der erste Prompt richtet sich innerhalb kollaborativer Reflexion an die Teilnehmer und fragt, was derjenige, der ein Problem thematisiert hat, hätte anders machen können. Hier wird von den Teilnehmern bereits eine kritische Auseinandersetzung mit dem Thema gefordert. Prompt 2 richtet sich ebenfalls an Teilnehmer kollaborativer Reflexion und fragt nach eigenen Erfahrungen und fokussiert somit den Teil des Erfahrungsaustausches in Reflexion. Ähnlich richtet sich der dritte Prompt ebenfalls an Teilnehmer kollaborativer Reflexion und fragt nach Lösungsvorschlägen. Die Zielsetzung ist hier Handlungsabsichten für die Zukunft zu entwickeln. Der vierte Prompt versucht allgemein Unerwartetes als Auslöser für

---

<sup>7</sup> Selbstgesteuertes Lernen beinhaltet unterschiedliche Aspekte: das selbstständige Auswählen und Nutzen von Lernstrategien, um ein gewähltes Ziel zu erreichen, das systematische Bewerten der eigenen Lerneffektivität und unabhängige Motivation insb. bei Erreichung von Lernzielen (Zimmerman 1990). Selbstwirksamkeit ist ein wichtiger Faktor beim selbstgesteuerten Lernen (Bandura 1977).

Reflexion (siehe Abschnitt 2.4) zu nutzen, um den Empfänger dazu zu bringen, über jüngere Erlebnisse zu reflektieren.

In der Reflexionsunterstützung können Prompts helfen, dass die Empfänger umfangreichere Beiträge schreiben, wenn nach Lösungsvorschlägen gefragt wird (Renner u. a. 2016). Prompts können gut eine Anleitung darstellen, um Personen beim reflexiven Schreiben (siehe Abschnitt 2.8.3) zu unterstützen (Lai und Calandra 2007; Roberts 2009; Loo und Thorpe 2002). Die Prompts müssen keine vollständige Anleitung sein, sondern können strukturgebend wirken (Winchester und Winchester 2011; Kori u. a. 2014; Loo und Thorpe 2002; Clarke 2004). Eine Frage wird häufig als Form von Prompts genutzt, um sowohl Reflexion als auch reflexives Schreiben zu unterstützen (Daudelin 1996; Loo und Thorpe 2002; Clarke 2004; van den Boom, Paas und van Merriënboer 2007).

Eine Studie, in der evaluiert wurde, wie gut Prompts Personen zum Treppensteigen ermutigen können, ergab, dass Prompts, die Konsequenzen (in dem Fall, falls die Person lieber den Aufzug nimmt) aufzeigen, effektiver sind als solche Prompts, die nur allgemein Informationen beinhalten. Es ist hilfreich, dem Empfänger mehrere unterschiedliche Prompts anzubieten, aus denen dieser auswählen kann (Yoo u. a. 2013; Saksono und Parker 2017; Scardamalia, Bereiter und Steinbach 1984).

Obwohl viele Studien zeigen, dass Prompts positive Effekte haben und die Empfänger zu einem bestimmten Verhalten motivieren können, ist dieser nachgewiesene Effekt teilweise sehr niedrig (Renner u. a. 2016).

Häufig werden Prompts als *Scaffolding* benutzt (Bannert 2009). Der Begriff *Scaffolding* kann einerseits als Substantiv verstanden werden, was ein Gerüst zur Unterstützung eines Lernalters meint (Pea 2004). Andererseits kann *Scaffolding* auch als Verb den Prozess des Unterstützens bezeichnen (Pea 2004). Wichtig sind zwei Aspekte bei Scaffolding: Erstens ist das Gerüst immer dazu da, den Lerner zu befähigen eine Aufgabe zu lösen, die er ohne dieses Gerüst nicht alleine hätte lösen können (Pea 2004). Zweitens ist dieses Gerüst nie von Dauer, sondern soll je nach Fortschritt des Lernalters sukzessive verringert bzw. entfernt werden (Pea 2004). Prompts eignen sich als Scaffolding, da sie, wie bereits beschrieben, den Lerner auf verschiedene Strategien aufmerksam machen, die dieser ggf. nicht spontan bereit hält (Deitz und Malone 1985; Gloger u. a. 2009; Thillmann u. a. 2009; Fivush, Gray und Fromhoff 1987). In der Literatur wird oft der zweite Aspekt von Scaffolding, dem Rückbau des Gerüsts, vernachlässigt (bspw. Mamykina u. a. 2008; McNicol u. a. 2014; Choi, Land und Turgeon 2008; Lee 2005), wobei Pea bemerkt hat, dass der Begriff häufig eher lose verwendet wird (Pea 2004). Es müsste sich über Zeit entweder der Inhalt der Prompts oder die Frequenz der Prompts ändern.

Zusätzlich muss die Frequenz der Prompts kontrolliert werden, da es zum Effekt des *Over-Prompting* kommen kann. In einer Studie von Nückles et al. zeigte sich, dass die Gruppe von Probanden, die Prompts für reflexives Schreiben (siehe Abschnitt 2.8.3) bekommen hat, sich zwar kurzfristig mehr Mühe beim Schreiben gegeben hat, aber zu Ende des Semesters weniger Mühe investiert hat als die Kontrollgruppe ohne Prompts (Nückles, Hübner und Renkl 2006). Dies deutet auf eine Art Ermüdungseffekt bei zu langem Prompting hin und Nückles empfiehlt, die Prompts ab einem bestimmten Zeitpunkt zu reduzieren (Nückles, Hübner und Renkl 2006). Dies wäre dann ein Ansatz, welcher der Beschreibung von Scaffolding am ehesten entspricht. Eine andere Studie bestätigt die Abnahme der Effektivität über Zeit (van Oldenbeek u. a. 2019). In einer anderen Studie, wo eine Gruppe viele sehr deutliche Prompts bekommen hat, hat diese Gruppe weniger gut abgeschnitten als die Gruppe, die keine Prompts bekommen hat (Holliday 1983). Die Studie behandelte allerdings nicht das Auslösen von Reflexion, sondern die Unterstützung von Wissen aus Erinnerungen im Bildungsbereich.

Vorteile von Prompts sind zum einen eine gute Möglichkeit der Skalierung, da sie überwiegend von einem Softwaresystem aus gesteuert werden (bspw. Davis und Linn 2000; Furberg 2009; Bannert 2006). In dem Fall wird keine menschliche Unterstützung benötigt, die in großen Gruppen schwer zu skalieren ist, bzw. generell viel Zeit in Anspruch nimmt (bspw. Morris und Stew 2007; van den Boom, Paas und van Merriënboer 2007).

Durch die gezielte Unterstützung einzelner Aktivitäten in einem Prozess (siehe activity Prompts in Tabelle 2-1), können Prompts sehr gut als adaptierbare Art der Unterstützung genutzt werden, weil sich die Texte je nach Kontext anpassen lassen. Dies ist allerdings nicht immer erfolgreich, wie die Studie bei Renner et al. zeigt, wo ein Prompt nach Lösungen gefragt hat, aber im Vergleich nicht mehr Lösungen aufgetaucht sind (Renner u. a. 2016).

Im Zuge der Anpassung ist es möglich, die Prompts direkt an den Lerner selbst anzupassen und so eine personalisierbare Form der Unterstützung zu erreichen. Die meisten der genannten Studien konzentrieren sich allerdings nur darauf zu untersuchen, ob Prompts wirken und welche Prompts zu welchem Ergebnis führen. Eine Form personalisierter Prompts, die sich auf den jeweiligen Nutzer selbst beziehen, findet sich in dem System von Kocielnik et al., wo das System versucht in einer Masse von Aktivitätsdaten Muster zu finden, um den Empfänger zu mehr physischer Aktivität zu

überzeugen. Dieses System ist allerdings eher ein Chatbot<sup>8</sup>, mit dem der Nutzer interagieren kann, als ein einfacher Prompt-Mechanismus.

Zusätzlich unterscheiden sich Personen nach ihrem Wissensstand. Experten haben im Vergleich zu Novizen bestimmte Schemata parat, die sie zur Problemlösung einsetzen können (vgl. für diesen Abschnitt Kalyuga 2009). Kalyuga et al. argumentieren, dass es sinnvoller ist, eine Unterstützung für Experten wegzulassen, bevor diese mit existierenden Schemata in Konflikt treten und damit einen negativen Effekt auf den Empfänger der Prompts haben kann. Dieser Konflikt kann auftreten, weil der Experte laut Kalyuga et al. vermutlich versuchen wird, die Hilfestellung mit seinem existierenden Wissen in Einklang zu bringen. Somit entstünde eine höhere kognitive Last für den Experten. Am meisten profitieren laut Kalyuga et al. Novizen von Hilfestellungen, wie bspw. Prompts und die Empfehlung ist, die Hilfestellung an den Lerner anzupassen (Kalyuga 2009). Somit können Prompts als Scaffolding genutzt werden, und über Zeit und bei hoher Expertise reduziert werden, um einerseits Over-Prompting zu vermeiden und andererseits die mögliche kognitive Last bei Experten zu reduzieren.

Nachteilig an Prompts ist, dass die Handlungsempfehlungen von einer Maschine kommen, die nach bestimmten Heuristiken handelt und somit häufig falschliegen kann, oder in manchen Situationen Prompts liefert, die nicht exakt zum Thema passen. Eine Maßnahme ist hier, dass der Nutzer aus unterschiedlichen Prompts wählen kann (Yoo u. a. 2013; Saksono und Parker 2017; Scardamalia, Bereiter und Steinbach 1984).

Zusätzlich kann Reflexion eine nachteilige Wirkung haben, wenn man zu lange über negative Dinge nachdenkt bzw. zu selbstkritisch ist (Verplanken u. a. 2007). Dies muss ein automatisiertes Prompting-System erkennen, um keine negative Wirkung auf den Nutzer auszuüben.

Das Entwerfen von Prompts ist eine ethische Frage, da der Entwickler von Prompts eine Handlungsabsicht des Empfängers beabsichtigt. Hier ist es wichtig ethisch korrekt zu handeln und den Nutzer nicht zu zweifelhaften Handlungen aufzufordern (Thaler und Sunstein 2008, 239–241).

Es ist festzustellen, dass zwar viele Paper eine Liste der verwendeten Prompts angeben (Renner u. a. 2016; Clarke 2004; van den Boom, Paas und van Merriënboer 2007; Thillmann u. a. 2009; Davis und Linn 2000; Davis 2003; Lin und Lehman 1999) aber

---

<sup>8</sup> Als Chatbots werden Programme bezeichnet, bei denen Menschen eine natürlichsprachliche Konversation mit einer Maschine führen können (Montero und Araki 2005). Häufig werden diese System für Kundenservice in Online-Shops genutzt (Montero und Araki 2005).

manche Studien keine oder nur ausgewählte Beispiele für die Prompts nennen (Bannert 2006; Bannert 2009; Lee 2005; Choi, Land und Turgeon 2008; Mamykina u. a. 2008; Furberg 2009).

### 2.10.3 Design von Prompts

Nachdem eine Definition für Prompts erarbeitet und die Wirkungsweise dargelegt wurde, wird an dieser Stelle eine Übersicht von aktuell in der Literatur bekannten Designempfehlungen gegeben.

Generell wird empfohlen, dass mit Prompts positive Verhaltensweisen ausgelöst werden, da der Empfänger sich besser fühlt und dann geneigter ist, diese auszuführen (McKenzie-Mohr 2000). Hinweise sind auch effektiv, wenn sie über mögliche negative Folgen der eigenen Handlung informieren (Thaler und Sunstein 2008, 36–37). Der Zeitpunkt eines Prompts ist relevant: Wenn der Empfänger die Prompts während eines spezifischen Teils einer Aktivität bekommt, ist es effektiver, als wenn die Prompts bspw. allgemein am Anfang geschickt werden (Thillmann u. a. 2009; Chung und Byrne 2008). Damit ist der Zeitpunkt gemeint, in dem es am wahrscheinlichsten ist, dass der Lernende Hilfe benötigt (Ifenthaler 2012). Im Umkehrschluss bedeutet das, dass man Prompts nicht schicken sollte, wenn man weiß, dass der Empfänger aktuell keine Zeit hat zu reagieren, etwa, weil eine andere Aufgabe dringlicher ist. Der Prompt selbst soll auch so gestaltet sein, dass er zum Handeln befähigt (Herrmann und Nierhoff 2017).

Wichtig ist, dass der Empfänger sieht, dass ein Prompt vorhanden ist, weil er sonst keine Möglichkeit hat zu reagieren (Butler und Winne 1995). Es hilft zusätzlich, wenn der Prompt aus den umgebenden Elementen hervorsticht, sprich, ein auffälligeres Design hat (McKenzie-Mohr 2000). Chung und Byrne nennen dieses Hervorstechen *saliency* und umfassen nicht nur visuelle Hinweise, sondern zusätzlich auditive, wie bspw. ein lauter Ton (Chung und Byrne 2008). Es bietet sich die Möglichkeit, die Prompts sehr auffällig zu gestalten, um in kritischen Moment die Aufmerksamkeit des Nutzers zu bekommen, bspw. über eine rot blinkende Fläche. Eine Empfehlung ist, den Prompt dort zu platzieren, wo der Nutzer seine Aktion durchführen soll (McKenzie-Mohr 2000). In einer Studie über verschiedene Möglichkeiten Aufmerksamkeit zu erzeugen, haben Gutwin et al. gezeigt, dass Elemente generell am sichtbarsten sind, wenn sie zentral im Sichtfeld eingeblendet werden (Gutwin, Cockburn und Coveney 2017). Manche Gestaltungsmerkmale, wie bspw. Farbe oder Form sind im peripheren Sichtbereich nicht mehr gut sichtbar, wohingegen sich bewegende, blinkende Elemente, oder Elemente



unterschiedlicher Helligkeit auch im Randbereich sehr gut Aufmerksamkeit auf sich lenken (Gutwin, Cockburn und Coveney 2017).

Bei Prompts ist es wichtig, eine passende Formulierung zu finden, damit der Empfänger den Prompt verstehen kann und somit eine Chance hat ihn befolgen zu können (Renner u. a. 2016). Idealerweise ist der Prompt selbsterklärend (McKenzie-Mohr 2000) und verwendet nur bekannte und eindeutige Begriffe (Schwarz und Oyserman 2001). Allgemein wird empfohlen, die Prompts so spezifisch wie möglich zu gestalten, um das gewünschte Verhalten beim Empfänger auslösen zu können (Glogger u. a. 2009; Chung und Byrne 2008; Davis und Linn 2000). Kocielnik et al. stellen fest, dass allgemeine Prompts dennoch wertvoll sein können (Kocielnik u. a. 2018). Eine Studie von Davis ergab, dass allgemeine Prompts zu einem besseren Gesamtverständnis führen und spezifische Prompts zur Reflexion stattdessen häufig zu unproduktiver Reflexion führen (Davis 2003). Allgemeine Prompts sind Prompts, die den Empfänger innehalten lassen und zum Nachdenken anregen, ohne eine Richtung anzugeben, worüber nachgedacht werden soll (Kori u. a. 2014).

Da ein Prompt nicht immer passend für den exakten Kontext des Nutzers sein kann, oder dieser über etwas anderes aktuell nachdenken möchte, kann es hilfreich sein mehrere unterschiedliche Prompts anzubieten (Yoo u. a. 2013; Saksono und Parker 2017).

Zur Unterstützung von Reflexion ist es wichtig, die Fragen nach *warum* vorsichtig zu formulieren, weil der Empfänger ansonsten defensiv reagieren kann (Vogt, Brown und Isaacs 2003). Zusätzlich wird empfohlen nach *was*, *wie* und *warum* zu fragen, und weniger nach *wer*, *wann*, *wo*, *welche* oder *ja/nein* (Vogt, Brown und Isaacs 2003). Dies wird in manchen Studien genutzt (King 1994; van den Boom, Paas und van Merriënboer 2007; Griffith und Frieden 2000). In anderen Studien werden häufig Fragen in Richtung gut verstandener bzw. noch nicht ausreichend verstandener Aspekte gestellt (Berthold, Nückles und Renkl 2007). Bei der Formulierung von Fragen sind drei Bestandteile wichtig: Das Fragewort selbst, der Umfang (im Orig. *scope*) und die Annahmen, die man über das Thema hat (Vogt, Brown und Isaacs 2003). Bei den Annahmen ist es wichtig, zu wissen, was der Empfänger bereits über das Thema weiß, um eine Erinnerung oder Aktivität möglichst passend auszulösen (Vogt, Brown und Isaacs 2003). Vogt et al. empfehlen auch, nicht davon auszugehen, dass etwas schief gelaufen ist, was mit der Vermeidung einer Defensivhaltung beim Empfänger einhergeht (Vogt, Brown und Isaacs 2003).

Wenn der Empfänger einen Prozess bearbeitet und an mehreren Stellen unterstützt werden soll, ist es wichtig, die Reihenfolge der Prompts zu beachten (Thillmann u. a.

2009). In einer Studie wurden beispielsweise zunächst Prompts zur Unterstützung der Informationssammlung gezeigt, bevor Prompts zur Verarbeitung dieser Informationen gezeigt wurden (Thillmann u. a. 2009).

Da Prompts idealerweise nur Erinnerung unterstützen und keine neuen Informationen beibringen (Deitz und Malone 1985; Glogger u. a. 2009), können beispielsweise Links in Prompts zu umfassenderen Informationen weiterleiten (Renner u. a. 2016).

Renner empfiehlt zur kognitiven Entlastung des Empfängers, dass Prompts erst nach einer Zeitspanne angezeigt werden, in der der Nutzer untätig war, um Hilfe anzubieten (Renner u. a. 2016). Auf diese Weise könne der Nutzer sich erst auf die Aufgabe konzentrieren, aber habe dennoch die Möglichkeit Hilfe zu erhalten, falls er nicht weiterwisse.

#### 2.10.4 Arten von Prompts

In der Literatur werden Prompts nicht nur als Prompts bezeichnet, sondern häufig auch noch spezialisiert verwendet, um ein konkretes Ziel zu erreichen. In Tabelle 2-1 werden unterschiedliche Arten von Prompts samt ihren Zielen aufgelistet. Da teilweise unterschiedliche Autoren synonyme Begriffe für die gleiche Art Prompt nutzen, werden diese entlang ihrer Bedeutung gruppiert.

**Tabelle 2-1: Unterschiedliche Arten von Prompts**

<b>Art des Prompts</b>	<b>Beschreibung</b>
<b>„Activity Prompt“ / Procedural Prompt / „Instructional Prompt“ / „Process Prompts“</b>	<b>Ziel:</b> Das Ziel dieser Art von Prompts ist, den Empfänger systematisch überprüfen zu lassen, ob dieser alle Teilaspekte einer Aktivität im Lernprozess ausreichend bedacht hat (Davis und Linn 2000; Scardamalia, Bereiter und Steinbach 1984; Bannert 2009; Lin u. a. 1999). Eine Studie ergab, dass Activity Prompts im Vgl. zu self-monitoring Prompts (s. u.) weniger gut geeignet sind, dass der Empfänger neues Wissen ihr in bestehendes integrieren können (Davis und Linn 2000). Bei Scardamalia et al. können die Empfänger aus einer Liste von Karten auswählen (1984). Das Ziel ist die Unterstützung von Strategien zur Erinnerung, und nicht um neues Wissen zu vermitteln (Bannert 2009). Bei Lin et al. werden diese Prompts als Möglichkeit dem Lernenden Fragen und Hinweise zum aktuellen Fortschritt zu geben, um den Lernprozess besser zu verstehen (1999). Dies wird bei Lin et al. als insbesondere hilfreich für Studenten beschrieben, die nach einem erfolglosen Versuch schnell sich einer anderen Lösung widmen,

Art des Prompts	Beschreibung
	<p>ohne vorher über das Scheitern der vorherigen Lösung nachzudenken (1999).</p> <p>Im weiteren Verlauf werden diese Prompts als Aktivitäts-Prompts bezeichnet.</p> <p><b>Beispiele:</b> „readers would get more from the article if the article...” (Davis und Linn 2000, 821) oder „An important point I haven’t considered is...” (Scardamalia, Bereiter und Steinbach 1984, 179)</p> <p><b>Von:</b> (Davis und Linn 2000; Scardamalia, Bereiter und Steinbach 1984; Bannert 2009; Lin u. a. 1999)</p>
<p><b>“Self-Monitoring Prompt”</b></p> <p><b>„Reflection Prompt“</b></p> <p><b>„Metacognitive Prompts“</b></p>	<p><b>Ziel:</b> Hier ist die Zielsetzung, beim Empfänger eine Reflexion über das eigene Lernverhalten auszulösen (Davis und Linn 2000). In der Studie bei Davis und Linn haben diese Prompts besser zur Reflexion beigetragen als Aktivitäts-Prompts (Davis und Linn 2000). Die self-monitoring Prompts werden von Davis weiterentwickelt und später als Reflection Prompt bezeichnet (Davis 2003). Dort unterscheidet diese zwischen directed Prompts und generic Prompts, die hinsichtlich der Spezifität verschieden sind. Operationalisiert werden diese Prompts dadurch, dass bei den allgemeinen Prompts abgefragt wird, worüber der Empfänger aktuell nachdenkt, und der zielgerichtete Prompt über einzelne Aspekte in einem Projekt reflektieren lässt (bspw. „To do a good job on our claim critiques, we need to...” (2003, 101). Die Studien haben gezeigt, dass diese Prompts besser zur Reflexion anregen als Aktivitäts-Prompts (Davis 2003; Davis und Linn 2000).</p> <p>Bei Bannert wird diese Art als <i>metacognitive Prompts</i> bezeichnet, womit diese die Unterstützung von Studenten bei der eigenen Informationsverarbeitung meint. Dazu werden Aktivitäten wie Orientierung, Zielsetzung, Planung, Beobachtung und Evaluation eingesetzt, um die Metakognition<sup>9</sup> und die Selbststeuerung zu fördern (Bannert 2009). Ziel ist die Förderung der Reflexion über das eigene Lernverhalten (Hoffman und Spataru 2008) und das Hinterfragen eigener Ansätze vor dem Hintergrund der gelernten Theorie (Fiorella und Mayer 2012).</p> <p>Im weiteren Verlauf werden diese Prompts als Reflexions-Prompts bezeichnet.</p>

<sup>9</sup> Metakognition beschreibt die Auseinandersetzung mit den eigenen Lernaktivitäten (wohingegen sich Lernaktivitäten auf ein Thema oder Gegenstand beziehen) (Molenaar, Slegers und Boxtel 2014).

Art des Prompts	Beschreibung
	<p><b>Beispiel:</b> „the part of critiquing that’s hardest for us is...” (Davis und Linn 2000, 821)</p> <p><b>Von:</b> (Davis und Linn 2000; Davis 2003; Bannert 2009; Hoffman und Spataru 2008; Fiorella und Mayer 2012; van den Boom, Paas und van Merriënboer 2007)</p>
<b>“Elaboration Prompt”</b> <b>„Problem Prompt“</b>	<p><b>Ziel:</b> Oft sind Erklärungen in der Reflexion nicht ausreichend oder oberflächlich. Daher verfolgen diese Art Prompts das Ziel, vom Empfänger umfangreichere Erklärungen zu bekommen (Ge und Land 2004; Fleck und Fitzpatrick 2010). In der kollaborativen Reflexion kann dies bspw. helfen, damit andere das Problem besser verstehen können. Man kann den <i>Elaboration Prompt</i> als Unterart der Aktivitäts-Prompts auffassen, da nur ein bestimmter Aspekt im Fokus ist.</p> <p><b>Beispiele:</b> „What is a new example of ...?“ oder „Why is it important?“ (Ge und Land 2004, 10); oder „What are the reasons this has gone badly/well?“ (Renner u. a. 2016, 3)</p> <p><b>Von:</b> (Ge und Land 2004; Fleck und Fitzpatrick 2010; Renner u. a. 2016)</p>
<b>Creativity Prompts</b>	<p><b>Ziel:</b> Prompts können in Kreativitätstechniken verwendet werden, um einzelne Aspekte verstärkt in den Fokus zu rücken, oder um komplexe Fragen bzw. Fragen mit mehreren Komponenten in einem Brainstorming in mehrere kleine Fragen aufzuteilen. Dies wird dann als <i>directed brainstorming</i> bezeichnet (Santanen, Briggs und de Vreede 2000).</p> <p><b>Beispiel:</b> Eine Brainstorming-Frage, wie bspw. „what do you think about air, water, ground pollution?“ wird in drei kleinere Fragen aufgeteilt: „what do you think about air pollution?“ und „...water pollution“ und „...ground pollution“ (Santanen, Briggs und de Vreede 2000, 4).</p> <p><b>Von:</b> (Santanen, Briggs und de Vreede 2000)</p>
<b>Novelty Prompts</b>	<p><b>Ziel:</b> In der Studie werden diese Prompts als <i>novelty messages</i> bzw. <i>novelty cues</i> bezeichnet, und sollen als Motivation für Freiwillige im Bereich <i>Citizen Science</i><sup>10</sup> dienen. Der Hintergrund ist, dass viele Personen motiviert sind, wenn sie etwas zuerst sehen bzw. bearbeiten</p>

<sup>10</sup> Als Citizen Science wird die Bestrebung bezeichnet, Bürger dichter an den Wissenschaftsprozess heranzuführen (Bonney u. a. 2016). Weiterhin wird darunter auch oft verstanden, dass Bürger Wissenschaftlern helfen, Daten zu sammeln (Bonney u. a. 2016).

Art des Prompts	Beschreibung
	<p>dürfen. Auf diese Weise soll die Beteiligung erhöht werden. Die Studie ergab, dass Personen dann pro Sitzung mehr beitragen und umfangreichere Beiträge leisten.</p> <p><b>Beispiel:</b> „Guess what! You’re the first to see this event“ (Jackson u. a. 2016, 174)</p> <p><b>Von:</b> (Jackson u. a. 2016)</p>
<b>Solution Prompts</b>	<p><b>Ziel:</b> Im Rahmen kollaborativer Reflexion ist ein Aspekt das Erarbeiten von möglichen Lösungen, die man danach in die Praxis überführen kann (siehe Abschnitt 2.5). Diese Art von Prompts soll die Teilnehmer in einer kollaborativen Reflexion dazu bewegen, Lösungsvorschläge zu einer Diskussion beizutragen. Dies hat in einer Studie dazu geführt, dass die Teilnehmer längere Beiträge geschrieben haben. Im Vergleich zu den übrigen Prompts in der Studie hat dieser Prompt aber nicht zu mehr Lösungsvorschlägen in den Antworten geführt! Von der Zielsetzung der Prompts her, passt dieser Prompt am ehesten zu der Art der Reflexions-Prompts, unterscheidet sich aber dahingehend, dass hier Lösungsvorschläge als spezifischer Aspekte im Fokus sind, während bei Reflexions-Prompts Reflexion allgemein im Vordergrund steht.</p> <p><b>Beispiel:</b> “Do you have an idea how to deal with this in the future?” (Renner u. a. 2016, 4)</p> <p><b>Von:</b> (Renner u. a. 2016)</p>
<b>„Content-oriented Prompts“</b>	<p><b>Ziel:</b> Als Ebene zwischen dem Anregen der Überprüfung eines Prozesses (Aktivitäts-Prompts) und dem Anregen der Reflexion selbst (Reflexions-Prompts) skizziert Furberg inhaltsorientierte Prompts, die im Bildungsbereich zur Überprüfung des Lernfortschritts dienen (Furberg 2009).</p> <p><b>Beispiel:</b> „Why is the order of the bases in a gene significant?“ (Furberg 2009, 405)</p> <p><b>Von:</b> (Furberg 2009)</p>
<b>„Point of Decision Prompts“</b> <b>„Point of Choice Prompts“</b>	<p><b>Ziel:</b> Die Zielsetzung ist, den Nutzer an einer Stelle, an der er eine Entscheidung treffen kann, zu einer bestimmten Entscheidung zu motivieren (Nocon u. a. 2010). Ein Beispiel dafür ist ein Schild an einem Lift oder Treppenaufgang, um den Nutzer dazu zu motivieren Treppen zu steigen. Zahlreiche Studien haben gezeigt, dass ein</p>

Art des Prompts	Beschreibung
	Hinweis auf die Gesundheit dazu führt, dass Personen eher die Treppen nutzen (Nocon u. a. 2010)
	<b>Beispiel:</b> „[...] regular stair climbing [...] provides daily exercise [...] [and] [...] keeps you fit [...]“. (Webb und Eves 2007, 51)
	<b>Von:</b> (Kaczynski, Wilhelm Stanis und Hipp 2014; Nocon u. a. 2010; Webb und Eves 2007)

Aus der Tabelle 2-1 lässt sich entnehmen, dass es im Bildungsbereich im Wesentlichen drei Arten von Prompts gibt, die zur Überprüfung einzelner spezifischer Lerninhalte dienen (*content-oriented Prompts*), zur Überprüfung, ob in einem Prozess alle Inhalte ausreichend bedacht wurden (*Aktivitäts-Prompts*) und zur Anregung von Reflexion über das Gelernte oder über die Lernstrategien selbst. Ähnliche Einteilungen (Prompts zu Prozessen, Erklärungen und Reflexion) finden sich in der Literatur (Papadopoulos u. a. 2009). *Solution Prompts* passen am ehesten als Element in die Kategorie der Reflexions-Prompts. *Elaboration Prompts* sind vielseitiger einsetzbar und können neben Reflexion als Form von content-oriented Prompts betrachtet werden. Die *creativity Prompts* können gezielt Kreativitätsprozesse unterstützen und sind so vielseitig einsetzbar: einerseits für Lösungsfindung in Reflexion (als Fusion mit Solution Prompts) und andererseits im Bildungsbereich zur Erleichterung der Identifikation neuer interessanter Ansätze zur Lösung eines Problems. Die *novelty Prompts* haben einen sehr engen Fokus und ihr Einsatz ist vermutlich eher auf Plattformen beschränkt, in denen sich Einzelne engagieren sollen und es Neues zu entdecken gibt.

Teilweise lassen sich nicht alle Studien in dieses Schema einsortieren, da zwar unterschiedliche Prompts genutzt werden, diese aber von der Formulierung her nicht eindeutig den obigen Kategorien zuzuordnen sind. Ein Beispiel dafür ist die Studie von Lin und Lehmann, die folgende Prompts wählen (Lin und Lehman 1999, 840):

- “reason justification (students were prompted to give reasons for their actions)”
- “rule based (students were prompted to explain rules or procedures)”
- “emotion focused (students were prompted to reflect on their feelings)”

Diese Prompts passen weniger zu den obigen Kategorien, da hier mehr eine inhaltliche Ausrichtung vorliegt. Aufgrund dieser Verteilung ist anzunehmen, dass *rule-based* Prompts vermutlich einen stark erklärenden Faktor haben und somit als *content-oriented* einzusortierenden sind. Ein Beispiel für rule-based Prompts ist “What conclusions can you draw from your experiments?” (Lin und Lehman 1999, 841), welches von der

Formulierung her einen stark reflexiven Charakter annimmt. Je nach Wahl der Prompts können *reason justification* Prompts eher reflexiv oder eher erklärend (*elaboration Prompt*) sein. Ein Beispiel für *reason justification* ist der Prompt “How did you decide that you have enough data to make conclusions?” (Lin und Lehman 1999, 841), welches von der Formulierung nicht eindeutig zuzuordnen ist und vom Empfänger sowohl als reflexiver als auch als erklärender Prompt interpretiert werden kann. Lin und Lehmann fassen zusammen, dass die *reason justification* Prompts zu dem größten Lernerfolg geführt haben (Lin und Lehman 1999).

### **2.10.5 Abgrenzung von anderen Begriffen**

Obwohl es unterschiedliche Begriffe gibt, die synonym zu Prompts verwendet werden, wie beispielsweise *guiding questions* (King 1994; King und Rosenshine 1993) oder ganz allgemein *cues* (Chung und Byrne 2008), gibt es ähnliche Begriffe, die sich inhaltlich von Prompts unterscheiden. In diesem Abschnitt wird eine Abgrenzung der Begriffe vorgenommen.

#### **2.10.5.1 Fragen**

Zunächst kann Prompting von einfachen Fragen abgegrenzt werden. Bei dieser Abgrenzung ist relevant, inwieweit die Fragen zur Reflexion genutzt werden. Wie zuvor gezeigt, sind Fragen auch eine häufige Form von textuellen Prompts (siehe Abschnitt 2.10).

Beispielsweise verwendet Daudelin Fragen zur Steuerung von Reflexion und greift dabei unterschiedliche Aspekte wie Problembeschreibung, Analyse eines Problems oder auch Planung einer Handlungsabsicht auf (Daudelin 1996). Bei Daudelin werden die Fragen in der Studie als Strukturierung der Aufgabe der Reflexion verwendet. Hier steht klar die Aufgabe der Reflexion im Vordergrund und die Fragen werden als Struktur und weniger als optionale Prompts verwendet.

An dem Beispiel wird deutlich, dass Fragen zwar Elemente von Prompts sein können, wie bspw. in der Literatur verwendet (bspw. Renner u. a. 2016; Bannert 2006; Kocielnik u. a. 2018), aber noch nicht eigenständig einen Prompt darstellen. Neben der Frage, muss auch ein Kontext vorhanden sein, in dem der Prompt dem Nutzer eine Handlungsoption aufzeigt, welche für diesen optional nutzbar ist.

### 2.10.5.2 Nudging

Nudges sind, nach dem gleichnamigen Buch von Thaler und Sunstein (2008), ebenfalls eine Möglichkeit jemanden zu beeinflussen. Der Begriff liegt inhaltlich dicht an Prompts dran und wird teilweise synonym verwendet (van Oldenbeek u. a. 2019), wobei es Aspekte gibt, bei denen sich die Konzepte voneinander unterscheiden.

Die Idee hinter Nudges ist, eine *choice architecture* aufzubauen, in welcher der Nutzer am ehesten die vom Architekten gewünschte Option auswählt, ohne Optionen zu verbieten oder ökonomische Anreize zu ändern (Thaler und Sunstein 2008, 6). Als einfaches Beispiel für eine Abgrenzung führen Thaler und Sunstein an, dass ein Nudge das Platzieren von Obst auf Augenhöhe ist, weil es dort am ehesten gesehen wird, und dass das Verbot von ungesunder Nahrung kein Nudge ist (Thaler und Sunstein 2008, 6). Ein weiteres Beispiel für Nudging ist die Benennung von Auswahloptionen in einem Dialog, über den man online an eine Organisation spenden kann (Thaler und Sunstein 2008, 24): Wenn höhere Beträge verwendet werden (bspw. 50 €/100 €/150 €/Mehr anstatt von 5 €/10 €/15 €/Mehr), dann werden im Durchschnitt höhere Beträge gespendet. Wie diese und andere Beispiele im Buch verdeutlichen, handelt Nudging überwiegend von der Platzierung von Auswahloptionen, die so angeordnet werden, dass die gewünschte Option wahrscheinlich ist. Im Unterschied dazu ist der Prompt eine explizite Aufforderung, bspw. in Form von Fragen oder Aufforderungen, selbst eine Handlung durchzuführen. Diese explizite Aufforderung zur Handlung fehlt jedoch bei Nudging. Zusätzlich bietet Nudging eher eine Lenkung des Nutzers zwischen Optionen an und Prompting versucht zu beeinflussen, wie eine Tätigkeit durchgeführt wird, bspw. durch Prompts über bestimmte Handlungen nachzudenken.

Nudging beruht auf dem Prinzip, dass Menschen ihre Entscheidungen aufgrund bestimmter Prinzipien treffen. Generell ist es so, dass die meisten Menschen mit der Auswahloption fortfahren, die als Standardoption gesetzt ist, bspw. beim Kauf von konfigurierbaren Gütern (Thaler und Sunstein 2008, 8). Diese Trägheit von der Standardoption abzuweichen macht sich Nudging zunutze (Thaler und Sunstein 2008, 8). Kognitiv gibt es zwei Systeme, die zur Entscheidungsfindung beitragen: ein automatisches System und ein reflexives System (vgl. für diesen Abschnitt Thaler und Sunstein 2008, 19–22). Ersteres trifft schnelle ad hoc Entscheidungen („Bauchgefühl“) und Letzteres ist für Entscheidungen, die auf Nachdenken basieren, verantwortlich. Oft komme es jedoch zu Fehlern, weil man zu oft auf sein Bauchgefühl hört und schnelle Entscheidungen trifft. Aus diesem Grund macht es Sinn, eine sinnvolle Option für schnelle Entscheidungen anzubieten. Eine *choice architecture* ist laut Thaler und Sunstein vorteilhaft, da nicht jeder ein Experte auf allen Gebieten ist (bspw. Ausgestaltung



betrieblicher Altersvorsorge), und es daher sinnvolle Auswahloptionen geben sollte. Insbesondere bei einer großen Anzahl von Auswahlmöglichkeiten kann man sich nicht mehr jede Option in der gebotenen Tiefe anschauen (Thaler und Sunstein 2008, 94). Sie empfehlen Nudges daher insbesondere für Entscheidungen, die schwierig und selten getroffen werden (Thaler und Sunstein 2008, 72). Eine umfangreiche Übersicht über unterschiedliche psychologische Wirkmechanismen bei der Entscheidungsfindung gibt es in Mirsch, Lehrer und Jung 2017; und Thaler und Sunstein 2008. Es gibt ebenfalls Konzepte, die abbilden, inwieweit diese Mechanismen in Software unterstützbar sind (Johnson u. a. 2012).

Beide Konzepte haben gemeinsam, dass sie den Empfänger nicht dazu zwingen, auf den Nudge (Thaler und Sunstein 2008, 6) oder Prompt (Fivush, Gray und Fromhoff 1987; Prilla 2015; Saksono und Parker 2017) zu reagieren. Thaler und Sunstein präzisieren, dass es leicht und günstig sein sollte den Nudge zu vermeiden (Thaler und Sunstein 2008, 6).

Im Bereich von Benutzerschnittstellen wird von *digital nudging* gesprochen (Weinmann, Schneider und vom Brocke 2016). Hier findet sich die obige Unterscheidung in einem Beispiel wieder: In einer Anwendung zum Bezahlen ist die Auswahl für Trinkgeld direkt ausgewählt und der Nutzer muss diese aktiv abwählen, wenn er kein Trinkgeld geben möchte (Weinmann, Schneider und vom Brocke 2016). Ein Prompt hätte an dieser Stelle gefragt, ob man zusätzlich ein Trinkgeld geben möchte. Hier zeigt sich also die Unterscheidung zwischen Nudge, dem Anbieten sinnvoller Standardoptionen, und einem Prompt, dem direkten Hinweisen auf eine Handlungsoption.

Bei Point of Decision Prompts (bspw. Nocon u. a. 2010; Kaczynski, Wilhelm Stanis und Hipp 2014; Webb und Eves 2007) ist der Unterschied zu Nudges geringer. Bei diesen Studien wird oft ein Schild aufgestellt, was die Vorzüge des Treppensteigens anpreist. Die Person muss sich entscheiden, welchen Weg sie gehen wird. Das Schild ist so gestaltet, dass eine der Möglichkeiten deutlicher hervorgehoben wird. Daher überschneiden sich die Konzepte Prompt und Nudge zum Großteil. Es lässt sich argumentieren, dass es sich um einen Prompt handelt, weil es einen aktiven Hinweis auf eine Handlung gibt und doch keine der Optionen physisch in den Vordergrund gerückt wird, um, wie bei Nudges, die naheliegende Option zu sein.

### **2.10.5.3 Notifikation**

Prompts haben den Charakter den Empfänger auf eine Handlungsoption hinzuweisen bzw. es wahrscheinlicher zu machen, dass der Empfänger auf eine bestimmte Art und Weise handelt (siehe Abschnitt 2.10). Eine Notifikation hingegen hat einen rein

informierenden Charakter und soll den Nutzer über etwas benachrichtigen (Palanque, Winckler und Martinie 2011). Dies wird oft in kollaborativen Systemen genutzt, um jemanden über die Änderungen aufmerksam zu machen (im Orig. *awareness*), die andere getätigt haben (Carroll u. a. 2003). Die Notifikation beabsichtigt nicht, dass der Empfänger tätig wird (Palanque, Winckler und Martinie 2011).

Beispiele für Notifikationen sind bspw. die Benachrichtigung über eine E-Mail oder der Hinweis, dass ein Kollege etwas in einem geteilten Dokument editiert hat (Carroll u. a. 2003). An dieser Stelle unterscheiden sich diese Notifikationen von Prompts dahingehend, dass diese zwar darauf aufmerksam machen, dass etwas passiert ist (Ankunft der E-Mail, Eingang einer Änderung), allerdings sind Notifikationen sprachlich so gestaltet, dass diese den Empfänger nicht zu einer Handlung auffordern. Prompts hingegen, wie bspw. „Have you been in a similar situation? What did you do?“ (Renner u. a. 2016), versuchen sprachlich deutlicher den Empfänger zur Beantwortung der Frage zu verleiten. Aus diesem Grund kann man zwar theoretisch zwischen Prompts und Notifikation unterscheiden, aber aufgrund der inhärenten Ungenauigkeit in der Sprache, kann die Ausprägung dieser Begriffe in der Realität sehr dicht beieinanderliegen.

## **2.11 Zusammenfassung**

Dieses Kapitel hat gezeigt, dass Reflexion eine Form informellen Lernens ist, die allein (siehe Abschnitt 2.4) und kollaborativ (siehe Abschnitt 2.5) durchgeführt wird. Reflexion umfasst das Evaluieren der eigenen Erfahrungen, mit der Zielsetzung Handlungsabsichten für die Zukunft zu entwickeln, um sich in ähnlichen Situationen anders verhalten zu können. Reflexion beinhaltet Lernen aus den eigenen Erfahrungen. Bei kollaborativer Reflexion können andere von den eigenen Erfahrungen lernen, man kann selbst aus den Erfahrungen Dritter lernen und gemeinsam Lösungsideen für künftige Handlungen entwickeln. Reflexion ist eine beliebte Form des Lernens im beruflichen Kontext (siehe Abschnitte 2.5.4 und 2.6). Ein Ort, an dem im beruflichen Kontext reflektiert wird, ist eine Community of Practice (siehe Abschnitt 2.1). Dabei handelt es sich um eine Gruppe von Personen, die eine gemeinsame berufliche Praxis hat und das Ziel verfolgt, weiter über die eigene Praxis zu lernen.

Dieses Kapitel hat ebenfalls gezeigt, dass (kollaborative) Reflexion in der Praxis unterschiedlichen Herausforderungen unterliegt, wie bspw. mangelnde Zeit (bspw. Renner u. a. 2016; Hatton und Smith 1995; Moon 1999, 166; Fleck und Fitzpatrick 2010), fehlende Gelegenheiten (Hatton und Smith 1995), häufige Ablenkung durch andere Tätigkeiten (Prilla, Degeling und Herrmann 2012), fehlende Strukturierung (Fleck und

Fitzpatrick 2009; Renner u. a. 2019), oder fehlende Motivation (Clarà 2015) (siehe Abschnitt 2.8). Dementsprechend ist es notwendig, dass Reflexion systematisch unterstützt werden muss, damit es in der Praxis funktioniert. Diese Unterstützung von Reflexion lässt sich auf unterschiedliche Arten gewährleisten, wozu bspw. das reflexive Schreiben in einem Journal oder die Arbeit mit Materialien in E-Portfolios gehört (siehe Abschnitt 2.8). Die Nutzung von E-Portfolios hat den Nachteil, dass sie zunächst nur die Sammlung von Materialien beinhaltet aber darüber Reflexion nicht direkt unterstützt wird. Dem Lernenden ist darüber weiterhin die Reflexion selbst überlassen oder wird durch Dritte, wie bspw. Lehrende, unterstützt. Bei Journalen ist dies ähnlich. Reflexives Schreiben in Journalen hilft bei der Reflexion, benötigt aber ebenfalls häufig weitere Unterstützung durch Dritte. Eine weitere Möglichkeit der Unterstützung von Reflexion ist die Nutzung von Prompts (siehe Abschnitt 2.10).

Prompts eignen sich aus unterschiedlichen Gründen zur Unterstützung von Reflexion. Prompts können automatisiert maschinell an den Empfänger verschickt werden, wodurch dies ein Mittel der Unterstützung ist, was auch für größere Zahlen von Nutzern skaliert. Gleichzeitig gibt es unterschiedliche Studien, die gezeigt haben, dass Prompts bei der Unterstützung individueller Reflexion helfen können (siehe Abschnitt 2.10). Prompts haben die Möglichkeit unterschiedliche Inhalte zu transportieren und können so versuchen unterschiedliche Aktionen beim Empfänger auszulösen. Aus diesem Grund können sich Prompts eignen unterschiedliche Aktivitäten von kollaborativer Reflexion, wie bspw. den Austausch von Erfahrungen oder das Entwickeln von Lösungen in einer gemeinsamen Diskussion, bspw. in einer Community of Practice Plattform zu unterstützen.

Die Grundlagen der Reflexion und der Prompts (siehe Abschnitt 2.10) werden im folgenden Kapitel in einem Konzept zusammengebracht. Dieses umfasst Prompts, die unterschiedliche Aspekte kollaborativer Reflexion systematisch durch Prompts anregen sollen, um eine Unterstützung von kollaborativer Reflexion zu erzielen. Um eine passende Unterstützung an die Theorie von kollaborativer Reflexion anbieten zu können, wird dazu die Basis in Abschnitt 2.9 genutzt, bei der unter anderem das CSRL-Modell für software-basierte Unterstützung von kollaborativer Reflexion beschrieben wird.



### **3 Konzept zur Unterstützung von Reflexionszyklen durch Prompts**

In diesem Kapitel wird ein Konzept beschrieben, welches eine systematische Unterstützung von Reflexion im Alltag durch Prompts gewährleisten soll. Dieses Konzept verbindet die unterschiedlichen darunterliegenden theoretischen Grundlagen, die in Kapitel 2 beschrieben wurden.

#### **3.1 Motivation und Ansatz**

Die vorherigen Abschnitte haben gezeigt, dass Reflexion ein anerkanntes und funktionierendes Werkzeug informellen und erfahrungsbasierten Lernens am Arbeitsplatz ist (siehe Abschnitt 2.4 und 2.5) und von Unternehmen aktiv genutzt wird (siehe Abschnitt 2.6). Personen reflektieren sowohl individuell als auch kollaborativ. Ein Ort, der zur kollaborativen Reflexion genutzt wird, ist eine Community of Practice (siehe Abschnitt 2.1). Innerhalb einer Community of Practice können sich Personen mit gleicher beruflicher Praxis über ihre Tätigkeiten und Erfahrungen austauschen. Insbesondere in verteilt arbeitenden Teams, Unternehmen mit mehreren Standorten oder in unternehmensübergreifenden Communities of Practice wird fachlicher Austausch betrieben und gemeinsam, bspw. über schwierige Herausforderungen in der Praxis, reflektiert. Aufgrund der zunehmenden globalen Vernetzung ist zu vermuten, dass diese Form der kollaborativen Reflexion zunimmt.

Eine Herausforderung bei der kollaborativen Reflexion ist, dass Unterstützung notwendig ist, damit Reflexion in der Praxis auch abläuft (siehe Abschnitt 2.8.1). Häufige Probleme sind fehlende Zeit (Renner u. a. 2016; Hatton und Smith 1995; Moon 1999, 165; Fleck und Fitzpatrick 2010), fehlendes Training (Platzer, Snelling und Blake 1997) oder auch fehlende Motivation (Clarà 2015). Aus diesem Grund gibt es unterschiedliche Unterstützungsformen für Reflexion, wie bspw. Journale (siehe Abschnitt 2.8.3), E-Portfolios (siehe Abschnitt 2.8.4), Visualisierungen (siehe Abschnitt 2.8.5) oder Prompts (siehe Abschnitt 2.8.6). Da bei Journalen und E-Portfolios weiterhin oft die Unterstützung bspw. durch Lehrende notwendig ist, skalieren diese Unterstützungsmechanismen nicht bei größeren Gruppen von Reflektierenden oder in Situationen, in denen spontan Unterstützung benötigt wird. Aus diesen Gründen konzentriert sich diese Arbeit auf Prompts, da diese unterschiedliche Vorteile haben. Im Gegensatz zu Unterstützung durch Kollegen oder Trainer können Prompts von einem System aus gesteuert werden und sind somit leichter skalierbar als menschliche Unterstützung. Zusätzlich sind Prompts der Weg, über den jemand zur Reflexion angeregt wird. Prompts können dementsprechend flexibel unterschiedlich dargestellt werden: bspw. visuell (McKenzie-Mohr 2000), textuell (Davis 2003; Thillmann u. a. 2009; Bannert 2006) oder auditiv (McKenzie-Mohr

2000). Bisherige Arbeiten zu Prompts haben gezeigt, dass Prompts helfen können Reflexion anzuregen (bspw. Renner u. a. 2016; Furberg 2009; Bannert 2006; Davis 2003). Bisherige Arbeiten fokussieren sich überwiegend auf individuelle Reflexion und haben kollaborative Reflexion außen vor gelassen (Renner u. a. 2016). Während allgemeine Inhalte von Prompts auch sinnvoll sind (Kocielnik u. a. 2018), verfolgen viele Prompts in bisherigen Studien das Anregen einzelner Prozessschritte (Davis 2003), wie bspw. die Betonung der Wichtigkeit einer sinnvollen Problembeschreibung (Fleck und Fitzpatrick 2010). Basierend auf der Literatur zeigt sich, dass Prompts helfen können, Aspekte von Reflexion anzuregen. Prompts, die individuelle Reflexion fördern, eignen sich aber nicht automatisch für kollaborative Reflexion (van den Boom u. a. 2004; van den Boom, Paas und van Merriënboer 2007). Bei der kollaborativen Reflexion kommt der Austausch von Erfahrungen und gemeinsame Diskussion bspw. zur Entwicklung von Lösungen hinzu (siehe Abschnitt 2.5). Dementsprechend ist aktuell unklar, inwieweit Prompts kollaborative Reflexion unterstützen können und welche Faktoren bei dieser Unterstützung zu beachten sind. Da bekannt ist, dass Personen teilweise Reflexion trotz Training nicht richtig anwenden können (Platzer, Snelling und Blake 1997), soll mit den Prompts nicht versucht werden, kollaborative Reflexion als solches anzuregen, sondern einzelne Bestandteile kollaborativer Reflexion. *Diese Arbeit untersucht, inwieweit Prompts in der Lage sind, Bestandteile von schriftlicher kollaborativer Reflexion, wie bspw. den Austausch von Erfahrungen oder das Erarbeiten von Lösungsvorschlägen in einer Community of Practice, zu unterstützen.*

In dieser Arbeit werden dafür Prompts als skalierbares, textbasiertes und systemgesteuertes Unterstützungsmittel betrachtet (siehe Abschnitt 2.10). Auf diese Weise soll Reflexion unterstützt werden. Dazu werden im folgenden Abschnitt (3.2) bestimmte Aspekte kollaborativer Reflexion und dazu passende Prompts, die eine Unterstützung ermöglichen können, identifiziert und beschrieben. Im Abschnitt 3.3 erfolgt dann eine Verbindung des Konzepts zu den einzelnen Studien dieser Arbeit. Da sich das Design der Prompts bzw. das Layout der Studien innerhalb der einzelnen Studien unterscheidet, wird dies in den jeweiligen Abschnitten innerhalb der Studien separat beschrieben.

Zur Analyse, inwieweit der Einsatz von Prompts erfolgreich war bestimmte Aspekte von kollaborativer Reflexion anzuregen, müssen die geschriebenen Texte derjenigen, welche die Prompts empfangen haben, analysiert werden. Die dafür notwendige Inhaltsanalyse wird in Abschnitt 3.4 beschrieben.

Für die vorliegende Arbeit werden für die Studien unterschiedliche Begrifflichkeiten für das Diskussionsforum genutzt, die im Folgenden voneinander abgegrenzt werden: Mit

*Forum* ist ein Bereich gemeint, der mehrere unterschiedliche Diskussionen zu einem bestimmten Thema beinhaltet und gruppiert. Ein *Thema* (im Engl. *Thread*) ist eine der Diskussionen, die aus einem initialen Beitrag besteht und keine, eine oder mehrere *Antworten* enthält.

### **3.2 Prompts als Unterstützungsmöglichkeiten für Aspekte kollaborativer Reflexion**

In diesem Abschnitt wird ein Konzept auf Basis wichtiger Aspekte kollaborativer Reflexion entwickelt. Das Konzept beschreibt, wie einzelne Aspekte kollaborativer Reflexion durch Prompts unterstützt werden könnten. Gleichzeitig werden unterschiedliche Barrieren identifiziert, die kollaborative Reflexion in der Praxis behindern können und daher ebenfalls ein Ansatzpunkt wären, bei dem Prompts helfen können.

Hinsichtlich der Ebenen der Reflexion, die in dieser Arbeit unterstützt werden sollen, ist das Ziel, die von Moon als informell bezeichnete Art der Reflexion (siehe Abschnitt 2.4.6), zu unterstützen (Moon 2006a). Moon unterscheidet die informelle Reflexion von der akademischen Reflexion, welche eher in einem universitären Kontext abläuft und einem strikteren Prozess unterliegt (Moon 2006a). Aufgrund der universitären Situation ist zu vermuten, dass eher Unterstützung durch Lehrende vorhanden ist. Dementsprechend ist bei der informellen Reflexion ein größerer Unterstützungsbedarf zu vermuten.

Das Konzept basiert sowohl auf einem Modell für kollaborative Reflexion als auf einem Schema für manuelle Inhaltsanalyse von Texten, die Reflexion enthalten. Die Beweggründe für die Auswahl werden im Folgenden erläutert.

Das Konzept soll sich an einem Modell für kollaborative Reflexion orientieren. Hier bietet sich die Nutzung des CSRL-Modells an (Krogstie, Prilla und Pammer 2013), da dieses im Vergleich zu anderen Modellen, wie bspw. dem beschriebenen Modell für kollaborative Reflexion (siehe Abschnitt 2.5.2) von Prilla et al. (2012), bereits auf die IT-Unterstützung von Reflexion ausgelegt ist. Das Modell bildet kollaborative Reflexion als zyklisches Modell mit vier Schritten ab, die jeweils Eingaben und Ausgaben haben (siehe für die folgenden Abschnitte ebenfalls Abbildung 2-4 in Abschnitt 2.9). Das Modell wurde entwickelt, um die Gestaltung von Reflexionsunterstützung anhand der unterschiedlichen Phasen von Reflexion zu ermöglichen. Der Vorteil des Modells liegt darin, dass mit den Eingaben in den einzelnen Schritten Anforderungen existieren, die über ein Softwaresystem geprüft werden kann. So lässt sich sukzessive der Prozess der

kollaborativen Reflexion unterstützen. Eine erste Version dieses Konzepts, welches für unterschiedliche Elemente des CSRL-Modells Ideen für Prompts geliefert hat, wurde bereits publiziert (Blunk und Prilla 2015). In dieser Arbeit wird das Konzept erweitert und aktualisiert. Dieses Konzept versucht, die unterschiedlichen Schritte kollaborativer Reflexion aus dem CSRL-Modell mit einzelnen Prompts zu unterstützen.

Zusätzlich zu den Schritten, die im CSRL-Modell beschrieben werden, wird ein Kodierschema zur Inhaltsanalyse von Texten mit Reflexion, an dem im Rahmen der Dissertation mitgewirkt wurde, im Konzept verwendet. Die Notwendigkeit von Inhaltsanalyse für Reflexion und das dazu benötigte Kodierschema wird im kommenden Abschnitt 3.4 näher erläutert. Das Kodierschema wurde ausgewählt, weil es aus einer Literaturanalyse hergeleitet ist und ein breites Spektrum von Elementen kollaborativer Reflexion abbildet. Gleichzeitig hat die Nutzung des Kodierschemas den Vorteil, dass die daraus entnommenen Aspekte kollaborativer Reflexion später anhand des Kodierschemas analysiert werden können.

Im Folgenden wird das Konzept beschrieben (vgl. für die folgenden Abschnitte Blunk und Prilla 2015). Es werden unterschiedliche **Ziele** identifiziert, die Elemente kollaborativer Reflexion sind oder kollaborative Reflexion direkt unterstützen können. Im Anschluss an die unterschiedlichen Ziele sind Prompt-Beispiele für die einzelnen Ziele tabellarisch dargestellt (siehe Tabelle 3-1 auf S. 86).

Die erste Phase besteht aus der Durchführung von Aktivitäten aus denen Informationen (im Orig. *data*) entstehen, die zur Reflexion genutzt werden können.

**Z1. Dokumentation von überraschenden bzw. unerwarteten Erfahrungen in der täglichen Arbeit:** Wie in Abschnitt 2.4 beschrieben, wird Reflexion ausgelöst, in dem eine Erfahrung anders erlebt wird, als sie erwartet wurde (Dewey 1933; Schön 1983; Boud, Keogh und Walker 1985). Bei Baumer wird dies als *breakdown* bezeichnet (Baumer 2015). Neben automatisch erfassten Daten, über die reflektiert werden kann (siehe Abschnitte 2.3 und 2.8), ist es notwendig, dass bestimmte Vorgänge manuell dokumentiert werden (Scott 2010; Prilla, Pammer und Krogstie 2013). Prompts für dieses Ziel von Reflexion versuchen, die Empfänger dazu zu bewegen, mit überraschenden bzw. unerwarteten Vorfällen verbundene Erfahrungen in ihrer Tätigkeit hinsichtlich der Reflexion aufzuschreiben.

Mithilfe der Daten aus der ersten Phase wird in der zweiten Phase die Reflexion initialisiert. Dies muss aber nicht geschehen, weil bspw. die Person nicht weiter über bestimmte überraschende Vorfälle nachdenkt.



- Z2. **Zielsetzung für die Reflexion festlegen:** Boud et al. beschreiben, dass Reflexion als Prozess zielgerichtet ist (siehe Abschnitt 2.4.4) und das Ergebnis oft eine Änderung in dem eigenen Verhalten ist (Boud, Keogh und Walker 1985). Dementsprechend kann Reflexion durch Prompts dahingehend unterstützt werden, dass diese dem Empfänger helfen eine Zielsetzung für die Diskussion festzulegen. Dadurch kann die Diskussion ebenfalls zielgerichtet ablaufen und der Reflektierende weiß, in welche Richtung er die Diskussion lenken möchte. Diese Zielsetzung ist im CSRL-Modell Teil des Rahmens für die Reflexion selbst (Krogstie, Prilla und Pammer 2013). Innerhalb des Systems wurde eine Möglichkeit geschaffen, dass der Nutzer ein Diskussionsziel oder eine Frage benennen kann, die dann für andere als Prompt hervorgehoben wird, was in dieser Arbeit als nutzergenerierte Prompts bezeichnet wird (siehe Abschnitt 3.5.3.2).
- Z3. **Andere Personen involvieren:** Um von den bisherigen individuellen Tätigkeiten in eine kollaborative Reflexion zu wechseln, können andere Personen wie bspw. Kollegen involviert werden. Dies hat den Vorteil, dass diese bei der Problemlösung helfen können, Feedback geben können und gemeinsam Lösungen entwickelt werden (Høyrup 2004; Prilla, Pammer und Krogstie 2013; Woerkom und Croon 2008). Die Idee ist, dass die Zielpersonen ermutigt werden, insbesondere in einer Community of Practice-Umgebung, weitere Personen in die Diskussion einzuladen, mit denen sie bisher nicht gearbeitet haben, um breiteres Feedback zu bekommen (Manouselis u. a. 2012). Prompts können helfen, die Empfänger explizit darum zu bitten, andere Personen in eine Diskussion einzuladen. Mit Empfehlungsalgorithmen könnten Personen bspw. auf Basis ihrer beruflichen Rolle vorgeschlagen werden.
- Z4. **Unterstützung individueller Reflexion bei Beschreibung des Problems:** Die Idee ist, dass die reflektierenden Personen nicht nur ein Problem beschreiben, sondern auch versuchen, eine erste Interpretation zu liefern, um von einer rein deskriptiven Ebene (siehe Abschnitt 2.4.6) auf eine höhere Ebene bzw. reflexive Ebene zu gelangen (Fleck und Fitzpatrick 2010). Hier können Prompts helfen bei der Formulierung eines neuen Themas das Nachdenken und Einbringen einer Interpretation für das Problem anzuregen.
- Z5. **Ausbau der Problembeschreibung:** Während das vorherige Ziel der Reflexion auf eine Interpretation abzielt, ist die Idee an dieser Stelle den Nutzer dazu anzuregen, die Problembeschreibung möglichst detailliert zu formulieren. Das Einbringen eigener Erklärungen oder Sichtweisen (Fleck und Fitzpatrick 2010) kann so erleichtert werden. Ebenfalls können Prompts an dieser Stelle die Personen über die

Wortlänge des Beitrags in Kenntnis setzen, um sehr lange Beiträge, die womöglich nur überflogen werden, oder sehr kurze Beiträge, die nicht alle Informationen enthalten, zu vermeiden (Wise, Zhao und Hausknecht 2013).

Die Ziele in diesem Schritt helfen dem nachfolgenden Schritt, die eigentliche kollaborative Reflexion, zu unterstützen. Das Element *Plan Session* aus dem CSRL-Modell wird nicht explizit aufgegriffen, da es sich um eine organisatorische Phase handelt, die eher in persönlicher kollaborativer Reflexion vor Ort relevant ist, aber bei der Betrachtung von kollaborativer Reflexion in einer Community of Practice Plattform weniger relevant ist. Ein Prompt könnte aber bspw. dazu genutzt werden, um das Planen eines Treffens von Angesicht zu Angesicht zu unterstützen. Da sich diese Arbeit auf schriftliche kollaborative Reflexion konzentriert wird dieser Aspekt im Konzept nicht berücksichtigt. Die Ziele Z4 und Z5 tauchen nicht im CSRL-Modell auf und sind aus der genannten Literatur hergeleitet und vermuten Unterstützungspotenzial durch Prompts.

In der kollaborativen Reflexion, welches der dritten Phase im CSRL-Modell entspricht, findet die eigentliche Diskussion und das Erarbeiten der Lösungen statt. Hierzu werden unterschiedliche Ziele unterstützt.

- Z6. Fokussierung auf das Beitragen eigener Erfahrungen:** Dieses Ziel schließt sich an den Teil „*Make related experiences available*“ aus dem CSRL-Modell an (Krogstie, Prilla und Pammer 2013). Die Idee ist hier, dass Personen eher eigene Erfahrungen einbringen sollen als gelerntes Wissen wiederzugeben, da Reflexion Lernen aus Erfahrung ist (siehe Abschnitte 2.4 und 2.5). Dies hilft bei organisationalem Lernen (Knipfer u. a. 2013). Dieses Ziel basiert auf Studien, die zeigen, dass das Teilen von Erfahrungen mit einer höheren Wahrscheinlichkeit für reflexive Ergebnisse in Computer-unterstützter kollaborativer Reflexion einhergeht (Prilla u. a. 2015). Prompts sollen hier unterstützen, dass Empfänger aus eigener Erfahrung argumentieren.
- Z7. Feedback einholen:** Das Ziel ist, dass in der kollaborativen Reflexion mehrere Personen nicht nur Erfahrungen austauschen (Z6), sondern sich zusätzlich gegenseitig Rückmeldung geben (Høyrup 2004). Dieses Ziel wird nicht direkt im CSRL-Modell beschrieben, allerdings taucht gezieltes Einholen von Feedback in den Beispielen des Artikels auf (Krogstie, Prilla und Pammer 2013). Andere Autoren beschreiben Feedback ebenfalls als Teil von kollaborativer Reflexion (Boud, Keogh und Walker 1985, 56) und als Teil von Lernen (Butler und Winne 1995). Prompts können hier den Empfänger ermuntern, die Kollegen nach Rückmeldung zu eigenen Handlungen zu fragen.

- Z8. **Einbringen von Lösungsvorschlägen:** Das Ziel von Reflexion ist die Entwicklung neuer Handlungsweisen, um Probleme in Zukunft anders anzugehen. In kollaborativer Reflexion können andere Personen helfen, indem sie Erfahrungen einbringen (Z6) oder indem sie Lösungsvorschläge einbringen. Prompts für dieses Ziel sollen bewirken, dass die Empfänger aus der eigenen Erfahrung heraus Vorschläge einbringen (Woerkom und Croon 2008; Raelin 2002; Dyke 2006) (anstatt bspw. unbegründete Hinweise zu geben).
- Z9. **Single Loop Learning:** Reflexion ist eine Form informellen Lernens, welche in dieser Arbeit unterstützt werden soll. Dies kann auf der Ebene passieren, dass Erkenntnisse aus der Lösung des konkreten Problems abgeleitet werden, was *single loop learning* genannt wird (Argyris und Schön 1996). Prompts können helfen, indem sie Empfänger danach fragen, was sie bereits aus der Reflexion gelernt haben. Häufig werden gelernte Dinge bzw. Ergebnisse von Reflexion nicht dokumentiert (Prilla, Degeling und Herrmann 2012). Dementsprechend kann der Prompt an dieser Stelle helfen. Dies hilft die Diskussion zu fokussieren. Die Dokumentation von Gelerntem ist interessant für Personen, die bei einer Online-Diskussion in einer Community of Practice nur lesend beiwohnen.
- Z10. **Double Loop Learning:** Im Vergleich zum single loop learning des vorherigen Ziels (Z9), steht bei *double loop learning* ein Lernen aus Verallgemeinerung der aktuellen Problemlösung auf andere Bereiche im Vordergrund (Argyris und Schön 1996). Prompts können versuchen, gezielt nach Verallgemeinerungen aus dem in der Reflexion diskutierten Thema zu fragen. Für Personen, die einer Diskussion nur lesend beiwohnen, ist es interessant, die Dokumentation des Gelernten zu lesen.

Reflexion soll idealerweise zielgerichtet ablaufen und darin münden, dass sich der Reflektierende eine neue bzw. angepasste Handlungsweise für die Zukunft überlegt (siehe 2.4). Die folgenden Ziele zielen darauf ab, die Ergebnisse aus den neuen Handlungsweisen zu dokumentieren.

- Z11. **Überprüfung, ob die Phase der Anwendung der Ergebnisse erreicht wurde:** Trotz Fortschritte im Bereich von Texterkennung und –Interpretation, ist es wichtig zu überprüfen, ob sich das Diskussionsthema in die Richtung entwickelt, die der Ersteller des Themas beabsichtigt hatte. Das Ziel ist, dass Prompts den Autor des Themas fragen, ob dieser bereits geplant hat, wie er das Problem in Zukunft angehen möchte, oder ob weitere Diskussion notwendig ist.
- Z12. **Planung der Anwendung der Reflexionsergebnisse:** Ein Ziel von Reflexion ist, dass eine Absicht für neue Handlungsweisen entsteht (Schön 1983; Boud, Keogh

und Walker 1985). Um diese leichter in die Praxis zu überführen, soll durch Prompts der Einsatz der neuen Handlungsweisen detaillierter geplant werden. Dies soll die Wahrscheinlichkeit erhöhen, dass die Handlungsweise nicht nur theoretisch überlegt wird, sondern auch aktiv eingesetzt wird. Dies kann dazu führen, dass die Personen nach den ersten Erfahrungen mit der neuen Handlungsweise in einen Zyklus individueller oder kollaborativer Reflexion übergeht.

- Z13. Anwendung der Ergebnisse aus vorherigen Reflexionszyklen prüfen:** Um die Anwendbarkeit der Ergebnisse aus einem vorherigen Reflexionszyklus zu prüfen, sollen Prompts eingesetzt werden. Dies kann dazu führen, dass die Personen nach den ersten Erfahrungen mit der neuen Handlungsweise in einen Zyklus individueller oder kollaborativer Reflexion übergeht.

Tabelle 3-1 zeigt eine Übersicht von Ideen für textuelle Prompts für jedes einzelne Ziel. Da diese Prompts Teil des Forschungsprojekts EmployID (siehe Abschnitt 4.2) waren, sind diese auf Englisch verfasst. Im weiteren Verlauf werden aus Gründen der Konsistenz ausschließlich die englischen Prompts verwendet. Diese Liste erweitert die bereits entwickelten Prompts aus (Blunk und Prilla 2015).

**Tabelle 3-1: Liste von Prompts zu den jeweiligen Zielen**

<b>Ziel</b>	<b>Mögliche Prompts</b>
<b>Z1: Dokumentation von überraschenden Erfahrungen in der täglichen Arbeit</b>	Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?
	What was unexpected for you in your daily work in the last 2 weeks? Why?
<b>Z2: Zielsetzung für die Reflexion festlegen</b>	Try to make clear what you want to know from your colleagues.
<b>Z3: Andere Personen involvieren</b>	Who of your colleagues might help you in this? Mention him in your post to notify him or her (@username).
<b>Z4: Unterstützung individueller Reflexion bei Beschreibung des Problems</b>	What do you think about what has happened? Do you already know how you want to approach this differently in future?
	Try to describe your point of view using a metaphor.
<b>Z5: Ausbau der Problembeschreibung</b>	What did you already try to approach the situation?
<b>Z6: Fokussierung auf erfahrungsbasierte Beiträge</b>	Do you agree / disagree with any of the existing comments? What would you suggest from your experience? Why?

<b>Ziel</b>	<b>Mögliche Prompts</b>
<b>Z7: Feedback einholen</b>	What do you think about how the person dealt with the situation?
<b>Z8: Einbringen von Lösungsvorschlägen</b>	What would you personally suggest from your experience? Why?
<b>Z9: Single Loop Learning</b>	What have you learned from this discussion so far?
<hr/>	
<b>Z10: Double Loop learning</b>	
<b>Z11: Überprüfung, ob die Phase der Anwendung der Ergebnisse erreicht wurde</b>	Do you already have an idea how you want to tackle similar situations in future? If not, how about phrasing a new question?
<b>Z12: Planung der Anwendung der Reflexionsergebnisse</b>	What specific steps do you want to take to approach similar situations in future?
	Do you already have an idea how you want to change your approach to the topic in future? If not, how about phrasing a new question?
<b>Z13: Anwendung der Ergebnisse aus vorherigen Reflexionszyklen prüfen</b>	Have you already made experiences after making changes in your approach? If yes, what are those experiences?

Der Prompt für Single und Double Loop Learning wurde in dieser Arbeit einheitlich gehalten, da das Ziel im Vordergrund stand, die Empfänger allgemein dazu zu ermuntern Gelerntes zu dokumentieren. Ein Prompt, der spezifisch nach aus einer Erfahrung verallgemeinertem Wissen (bei Double Loop Learning) fragt, könnte für den Empfänger in einer Situation, in der nur kurz ein Beitrag in einer Diskussion verfasst werden soll, zu viel Aufwand verursachen.

Studien zeigen, dass die Reihenfolge der Prompts wichtig ist (Thillmann u. a. 2009). Viele der Ziele lassen sich jederzeit von Prompts unterstützen, und andere Ziele, wie bspw. Z9-Z13, die nach Gelerntem oder nach der Anwendung von neuen Lösungen fragen, sind erst sinnvoll, wenn die Diskussion eine bestimmte Länge erreicht hat. Eine Beschreibung dieses Aufbaus findet sich in der entsprechenden Studie (siehe Abschnitt 4.3).

Im folgenden Abschnitt (3.3) werden die Ziele, die in dieser Arbeit untersucht werden sollen, ausgewählt. Zusätzlich werden die Studien zur Überprüfung, inwieweit Prompts diese Ziele unterstützen können, beschrieben.

### 3.3 Studien

Nachdem in dem vorherigen Abschnitt verschiedene Ziele von Reflexion, die sich durch Prompts unterstützen lassen, aufgelistet wurden, wird in diesem Abschnitt beschrieben, welche Ziele konkret in dieser Arbeit in Studien untersucht werden sollen.

*Das Ziel der Arbeit ist die Analyse, inwieweit Prompts Aspekte von Reflexion in der täglichen Praxis in einer Community of Practice unterstützen können.* Dazu werden unterschiedliche Studie aufgesetzt.

In einer Feldstudie (siehe Kapitel 4) wurden die im Konzept herausgearbeiteten Ziele und die damit verbundenen Prompts aus Tabelle 3-1 getestet. Die Studie soll einerseits darüber Aufschluss geben, inwieweit Prompts bestimmte Aspekte von kollaborativer Reflexion anregen können und andererseits, inwieweit Prompts helfen können kollaborative Reflexion in der Praxis zu unterstützen. In der Studie wird das Ziel Z3 nicht betrachtet, da die für dieses Ziel notwendige Funktionalität in der Community of Practice Plattform nicht vorhanden war. Die Studie selbst findet in einer Community of Practice Umgebung statt, in der Nutzer sich über die tägliche Arbeit austauschen und gemeinsam voneinander lernen können. Prompts werden in der Studie eingesetzt, um Aspekte von Reflexion anzuregen und dadurch Reflexion wahrscheinlicher zu machen. Somit wird versucht, das Konzept direkt in einer Feldstudie zur Beantwortung der Forschungsfrage umzusetzen. Die Durchführung einer Feldstudie hat den Vorteil, dass die Prompts in einer für die Nutzer gängigen Umgebung eingesetzt werden. Aufgrund der Natur einer offenen Feldstudie gibt es viele unterschiedliche Einflüsse, die sich nicht kontrollieren lassen: bspw. ändert sich kontinuierlich die Zusammensetzung der diskutierenden Personen, es gibt neue Richtlinien in der Organisation, welche die Arbeit beeinflussen können, etc.. Aus diesem Grund werden zusätzliche Studien durchgeführt.

Neben der Feldstudie wird eine zweite Studie, die aus drei Teilstudien besteht, in einer Crowdsourcing-Umgebung durchgeführt. Eine Übersicht über Crowdsourcing als Umgebung für Studien findet sich in Abschnitt 3.5. Im Gegensatz zur Feldstudie hat diese Studie den Vorteil, dass die Umgebung etwas kontrollierter ist. Es wird in den Studien eine Forenumgebung simuliert, in der Probanden ein für sie relevantes Thema diskutieren. Da das Thema in der Studie für jeden Teilnehmer das Gleiche ist, ist diese Studie im Vergleich zu der Feldstudie mit wechselnden Themengebieten etwas kontrollierter. Weitere Vorteile einer Crowdsourcing-Studie sind, dass sich leicht eine hohe Anzahl von Probanden rekrutieren lässt (siehe Abschnitt 3.6.1) und die Ergebnisse, mit denen von Laborstudien vergleichbar sind (siehe Abschnitt 3.6.6). Hier kann gezielt die Wirkung von spezifischen Prompts überprüft werden.

In der Teilstudie 2.1 sollen zunächst zwei Kernelemente von kollaborativer Reflexion abgebildet und untersucht werden. Spezifisch soll die Studie darüber Aufschluss geben, *inwieweit sich Erfahrungen und Lösungsvorschläge durch Prompts in einer simulierten Community of Practice Umgebung anregen lassen* (siehe Abschnitt 5.1).

Bei der Gestaltung von Prompts ist ebenfalls relevant, dass der Prompt verständlich ist (siehe Abschnitt 2.10.3). Während die Prompts in Teilstudie 2.1 von der Formulierung her sehr unterschiedlich sind, sollen in Teilstudie 2.2 Prompts getestet werden, die hinsichtlich der Formulierung näher beieinanderliegen, aber doch unterschiedlich sind. Diese Studie evaluiert, *inwieweit feinere Nuancen in den Formulierungen von Prompts eine Rolle beim Anregen von Lösungsvorschlägen für kollaborative Reflexion spielen* (siehe Abschnitt 5.2).

Die gemeinsame Diskussion ist ein wichtiger Bestandteil kollaborativer Reflexion. In der Teilstudie 2.1 beantworten Teilnehmer ein neues Thema, welches keine Diskussionsbeiträge enthält. In Teilstudie 2.2 hat das gleiche Thema nun vordefinierte Antworten enthalten. In beiden Teilstudien sollen die Teilnehmer eine eigene Antwort beitragen. In beiden Studien wird zum Anregen von Lösungsvorschlägen derselbe Lösungs-Prompt verwendet. Die Frage ist hier, *inwieweit das Vorhandensein von Diskussionsbeiträgen einen Einfluss auf die Erstellung von Lösungsvorschlägen in einer simulierten Community of Practice Plattform hat*.

Neben der Anzahl entstandener Lösungsvorschläge durch Prompts in den Teilstudien 2.1 und 2.2 sollen diese Lösungsvorschläge ebenfalls qualitativ bewertet werden, um eine differenziertere Analyse zu ermöglichen. Dazu wird eine dritte Teilstudie durchgeführt. Da Reflexion auf Neubewertung und Austausch von Erfahrungen basiert (siehe Abschnitte 2.4 und 2.5), soll hier evaluiert werden, welche Qualität die Lösungsvorschläge aus den vorherigen Teilstudien haben. Dazu werden in der Teilstudie unterschiedliche Kriterien zur Bewertung erarbeitet. Die Studie soll die Qualität der Lösungsvorschläge zwischen den Teilstudien 2.1 und 2.2 analysieren und klären, *inwieweit sich die Qualität von Lösungsvorschlägen unterscheidet, wenn Lösungs-Prompts in Diskussionen mit und ohne Diskussionsbeiträgen genutzt werden* (siehe Abschnitt 5.3).

Diese zwei Studien sollen zur Beantwortung der Forschungsfrage beitragen und sowohl in einer Feldstudie als auch in einer simulierten Community of Practice Umgebung (innerhalb einer Crowdsourcing-Studie) analysieren, inwieweit Prompts helfen können Aspekte kollaborativer Reflexion anzuregen.

### **3.4 Qualitative Inhaltsanalyse zur Überprüfung der Prompts**

Diese Arbeit fokussiert sich auf schriftliche Reflexion in einer Community of Practice in der beruflichen Praxis und sie muss dementsprechend dort beobachtet und analysiert werden (Schön 1983, 169). Eine gängige Methode zur Analyse von Texten ist der qualitative Ansatz der Inhaltsanalyse. Dieser Ansatz ist bei der Analyse von Reflexion beliebt (bspw. Poldner u. a. 2014; Chirema 2007; Isaacs u. a. 2013) und allgemein bei der Analyse von online Diskussionen (bspw. Wever u. a. 2006; Hara, Bonk und Angeli 2000).

Es gibt unterschiedliche Arten der Inhaltsanalyse: Einerseits gibt es manuelle Inhaltsanalysen, in denen Forscher Texte analysieren und bestimmte Abschnitte manuell markieren. Automatisierte Inhaltsanalyse basiert entweder auf einem Wörterbuch mit festgelegten Stichworten je Kategorie oder auf maschinellem Lernen, welches oft auf manuell kodierten Texten basiert (Ullmann 2015b). Ein Beispiel für stichwortbasierte Systeme sind die Tools LIWC (Pennebaker u. a. 2015) und Empath (Fast, Chen und Bernstein 2016). Die Herausforderung bei der Nutzung von automatischer Inhaltsanalyse auf Basis von Wörterbüchern oder maschinellem Lernen ist, dass diese Systeme erst mit echten Daten trainiert bzw. kalibriert werden müssen. In zwei parallel zur Dissertation betreuten Abschlussarbeiten wurde versucht, eine Möglichkeit zur Inhaltsanalyse von Reflexion auf Basis des Kodierschemas von Prilla et al. zu schaffen (Prilla u. a. 2015). In beiden Arbeiten konnten keine zufriedenstellenden Ergebnisse erzielt werden, mit der die Inhaltsanalyse hätte automatisiert werden können. Die in den Systemen, bspw. in LIWC oder Empath, verfügbaren Kategorien sind für eine spezifische Analyse von Reflexion zu allgemein gehalten und nicht vergleichbar mit einem Schema, dass für eine detaillierte Inhaltsanalyse für Reflexion entwickelt wurde. Das im Folgenden verwendete Schema wurde für kollaborative Reflexion entwickelt.

#### **3.4.1 Kodierschema für Reflexion**

In dieser Arbeit wird das auf Reflexion zugeschnittene Kodierschema von Prilla et al. verwendet (Prilla u. a. 2015). In Tabelle 3-2 wird eine Übersicht über das Kodierschema gegeben. Dieses Schema wurde ausgewählt, da es auf einer Literaturanalyse unterschiedlicher Merkmale von kollaborativer Reflexion basiert. Dementsprechend eignet es sich zur Analyse der geplanten Studien. Nachteile bei der Nutzung dieses Schema ist die notwendige manuelle Inhaltsanalyse, da bisher keine automatisierte Möglichkeit existiert, das Schema auf Texte anzuwenden.



**Tabelle 3-2: Reflexionsbezogenes Kodierschema (angepasst aus (Prilla u. a. 2015);  
Beschreibungen wurden übersetzt und Beispiele sind übernommen.)**

<b>Code</b>	<b>Beschreibung</b>
<b>EMO_OWN</b>	Erwähnen eigener Emotionen innerhalb des Textes (bspw. „Was not fun man“ oder „this made me angry“). Es muss ein Wort vorhanden sein, was die Emotion explizit deutlich macht: bspw. fröhlich, traurig, frustriert oder wütend.
<b>EMO_OTH</b>	Es wird über die Emotionen anderer gesprochen (bspw. „[resident] said he is unhappy living here“). Es muss ein Wort vorhanden sein, dass die Emotion des anderen explizit macht: fröhlich, traurig, frustriert oder wütend.
<b>INT</b>	Interpretation oder Begründung von Aktionen und Situationen (bspw. „As far as I am aware I had done nothing to deserve this“). Der Code wird vergeben, wenn der Autor einer Diskussion eigene Interpretationen oder Begründungen hinzufügt, bspw. warum eine Situation problematisch oder relevant war und mögliche Hypothesen für Erfolg/Misserfolg einschließt. Dies geht dann über eine einfache Beschreibung der Situation hinaus.
<b>EXP</b>	Verbinden der Diskussion zu einer eigenen Erfahrung (bspw. „I made a similar experience “ oder „X told me he was through this before ...“). Es muss eine explizite Referenz auf die eigene Erfahrung vorhanden sein. Andernfalls wird ein anderer Code genutzt (bspw. KNO).
<b>KNO</b>	Verbinden der Diskussion zu eigenem Wissen, Regeln oder Werten, die nicht explizit mit einer eigenen Erfahrung verbunden sind (bspw. „I was reading an article (guardian newspaper) on do's and don'ts in an interview. One of a don't was, not to keep looking in the interviewees eyes but look at their face“).
<b>DISAGR</b>	Eine Antwort in einer Diskussion, bei der aktuelle Interpretationen oder Vorschläge abgelehnt werden, oder eigene Perspektiven hinzugefügt werden (bspw. „Hmmm. Is this really different from [...]“).
<b>AGR</b>	Eine Antwort in einer Diskussion, bei der aktuellen Interpretationen oder Vorschlägen zugestimmt wird (bspw. „Agreed!“). Der Code wird vergeben, wenn eine Antwort in der Diskussion vorherige Vorschläge wiederholt.
<b>ADV</b>	Ein Hinweis, der ohne eine Begründung gegeben wird (bspw. „never accept blame for another's mistake“).
<b>SUG</b>	Wenn ein Lösungsvorschlag mit einer Begründung gegeben wird. Dieser Code wird berechnet und ist vorhanden, sobald entweder ein Vorschlag auf Erfahrungsbasis (SUG_EXP) oder Wissensbasis (SUG_KNO) vergeben wird.
<b>SUG_ EXP</b>	Wenn ein Lösungsvorschlag mit einer Begründung gegeben wird und die Basis für diese Begründung eine eigene Erfahrung ist. Der Teil des Textes, der die Erfahrung für diesen Lösungsvorschlag repräsentiert, wird dann zusätzlich mit EXP (siehe oben) markiert.
<b>SUG_ KNO</b>	Wenn ein Lösungsvorschlag mit eigenem Wissen begründet wird. Diese Begründung wird im Text zusätzlich mit dem Code KNO markiert.
<b>S_LOOP</b>	Dieser Code bildet eigene Einsichten als single-loop learning ab, wenn diese aus der reflexiven Diskussion resultieren (bspw. „I realised that I shouldn't

Code	Beschreibung
	have have been so worried about this [...]“ oder „It is good to know that I personally haven’t done something wrong“). Der Code wird nur vergeben, wenn aus der Diskussion offensichtlich ist, dass die Person aus der Diskussion gelernt hat. Aussagen über Dinge, die außerhalb der Diskussion gelernt wurden, werden nicht kodiert.
<b>D_LOOP</b>	Dieser Code bildet Gelerntes in Form von double-loop learning ab: Eine Verallgemeinerung der Erfahrungen. Dies wird oft als Muster oder Ursprung eines Problems bezeichnet (bspw. „the best way I found to deal with this is [...]“). Der Code wird nur vergeben, wenn aus der Diskussion offensichtlich ist, dass die Person aus der Diskussion gelernt hat. Aussagen über Dinge, die außerhalb gelernt wurden, werden nicht kodiert.
<b>CHANGE</b>	Dieser Code repräsentiert den Versuch, Schlussfolgerungen aus der Reflexion in die eigene Praxis zu überprüfen. Dies umfasst bspw. neue Ansätze zur Problemlösung, Berichte über geplante oder bereits erfolgte Handlungsänderungen (auf Basis der Diskussion). Bspw. „Will definitely try and do [...] in the future“.
<b>Q_INF</b>	Fragen nach mehr Informationen über eine Situation. Diese Frage taucht auf, falls nicht alle Informationen gegeben wurden oder falls bestimmte Informationen verfügbar sein sollten. Bspw. „what do you mean by [...]“ oder „what happened afterwards“.
<b>Q_INT</b>	Fragen nach einer Interpretation, nach Meinungen oder nach Lösungsvorschlägen (bspw. „what do people think about [...]“). Dieser Code wird ebenfalls genutzt, wenn Fragen dazu dienen eine Diskussion zu starten oder fortzusetzen (bspw. „what do people think about [...]“ oder „how can we solve this problem?“).

In der Fassung in Prilla et al. verfügt das Kodierschema über eine Unterscheidung zwischen Vorschlägen mit (SUG) und ohne Begründung (ADV). Bei der Anwendung in der Praxis hat sich jedoch gezeigt, dass begründete Lösungsvorschläge oft einen erfahrungsbasierten oder wissensbasierten Charakter haben. Daher wird in dieser Arbeit das Kodierschema, um zwei neue Codes erweitert, um sowohl erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) als auch wissensbasierte Lösungsvorschläge (SUG\_KNO) abbilden zu können. Diese beiden Codes können als Spezialisierung des Codes für allgemeine Lösungsvorschläge (SUG) verstanden werden. Umgekehrt betrachtet ist der Code SUG eine Verallgemeinerung der Codes SUG\_EXP und SUG\_KNO.

### **3.4.2 Erweiterung des Kodierschemas für Prompts**

Das reflexionsbezogene Kodierschema wurde um drei Codes erweitert, die abbilden sollen, wie sehr eine Antwort auf einen Prompt Bezug nimmt. Die folgenden Codes wurden neu hinzugefügt:

1. User did NOT respond to prompt.
2. User responded in PARTS to the prompt.
3. User responded fully to the prompt.

Diese Codes sollen einen Indikator geben, ob der Verfasser den Beitrag verfasst, ohne auf den Prompt einzugehen, nur teilweise auf den Prompt eingeht, oder ob der Text voll zu dem Prompt passt. Eine Restriktion bei der Anwendung der neuen Prompt-bezogenen Codes war, dass je Beitrag mit Prompt-Hintergrund maximal einer der Codes vergeben werden darf. Es ist zu unterstreichen, dass diese Form der Inhaltsanalyse nur besagt, dass eine inhaltliche Übereinstimmung vorhanden war, aber nicht, ob der Autor dies absichtlich getan hat. Falls der Autor des Themas beim Lesen des Beitrags an eine bestimmte Antwort gedacht hat, und der Prompt zufällig das gleiche Thema anspricht (bspw. das Einbringen eigener Erfahrungen), dann wird der Text dahingehend markiert, dass die Antwort mit dem Prompt übereinstimmt. Es ist daher möglich, dass diese Übereinstimmung zufällig entstanden ist. Dies lässt sich allerdings post-hoc aus den Daten nicht analysieren und es ist zu bezweifeln, dass die Nutzer dies nach mehreren Monaten noch wahrheitsgetreu in bspw. einem Interview oder einer Umfrage wiedergeben könnten. Für die Analyse stellt dieses Vorgehen jedoch eine angemessene Näherung dar.

### **3.4.3 Anwendung des Kodierschemas**

Für die Anwendung des Kodierschemas wurden ein paar Regeln erstellt, um das Kodieren zwischen mehreren Anwendern vergleichbar zu gestalten.

Kodiert werden in dieser Arbeit unter anderem Beiträge in Online-Diskussionen (in Studie 1, siehe Kapitel 4) und Antworten in simulierten Online-Foren (in den Studien 2, 3 und 4, siehe Abschnitte 5.1, 5.2 und 5.3). Es wird ein Beitrag, sei es der initiale Beitrag, der ein Thema beginnt oder die Antwort auf ein Thema, immer als ein Block interpretiert. Für die Analyse wird später berechnet, ob dieser Beitrag einen Code enthält. Die Analyse ist binär und es wird nicht die Anzahl der Codes gezählt. Die Einheit der Kodierung sind immer logische Einheiten. Dementsprechend kann ein Code auf mehrere Sätze

angewendet werden, die bspw. eine Erfahrung beschreiben, oder auf Satzteile, oder auf einzelne Wörter (bspw. Codes für Emotionen). Die Codes sind nicht exklusiv angelegt und daher können einzelne Satzteile bzw. Sätze mehrere Codes beinhalten. Dies ist beispielsweise der Fall, wenn jemand aus eigener Erfahrung berichtet und eigene Emotionen nennt. Dann ist der Teil des Textes sowohl mit Code EXP als auch mit Code EMO\_OWN belegt.

Wenn der Text kodiert wird, soll möglichst exakt die Textstelle markiert werden, auf die sich der Code bezieht. Wenn eine Person in einem Text mehrere Sätze schreibt, aber davon nur ein Satz eine Erfahrung beschreibt, dann soll auch nur dieser Satz mit dem Code für Erfahrungen markiert werden. Die übrigen Sätze werden nicht markiert. Dadurch wird die Kodierung insgesamt präziser.

Bei der Anwendung des Schemas ist die Regel, dass Textstellen nur dann mit einem Code versehen werden, wenn der Code innerhalb der Textstelle auf eine Phrase oder ein Wort bezogen werden kann. Anders formuliert: Der Text darf nicht frei interpretiert werden. Bspw. darf ein Text nur mit dem Code für Emotionen versehen werden, wenn ein Wort existiert, welches die Emotion ausdrückt (bspw. „wütend“ oder „frustriert“). Falls der Text nur so klingt, als sei der Verfasser fröhlich, dann wäre dies eine Interpretation und wird daher nicht kodiert. Damit soll die Interpretation eingeschränkt werden, um die Präzision der Kodierung zu erhöhen.

Bei der Anwendung des Kodierschemas sollen spezifische Codes bevorzugt vor abstrakteren Codes verwendet werden. Dies bedeutet, dass ein erfahrungsbasierter Lösungsvorschlag als SUG\_EXP kodiert werden soll und nicht nur als Lösungsvorschlag (SUG). Dies hilft ebenfalls präzisere Kodierungen zu erreichen.

### **3.4.4 Vorgehen bei der Inhaltsanalyse**

Das Vorgehen in dieser Arbeit ist angelehnt an das beschriebene Vorgehen aus der Quelle des Schemas (Prilla u. a. 2015).

Um eine objektive Analyse der Ergebnisse zu ermöglichen, werden die Textstellen in dieser Arbeit von mindestens zwei Personen kodiert. Jede Person kodiert unabhängig von der anderen Person alle Textstellen.

Das Vorgehen wurde so gestaltet, dass zunächst jede Person eine Teilmenge der Texte kodiert hat (ca. 20 Texte). Anschließend wurden die Kodierungen für diese Textstellen verglichen und bei Bedarf angepasst. Der Hintergrund ist, dass auf diese Weise beide

Personen den Umgang mit dem Kodierschema üben konnten. In der Praxis hat sich gezeigt, dass die Anwendung eines solchen Kodierschemas sehr anfällig für Flüchtigkeitsfehler ist, weil es leicht ist, bestimmte Aspekte in einem Satz zu überlesen. Anschließend wird die gesamte Menge an Beiträgen separat kodiert.

Es folgt eine Berechnung der Übereinstimmung der beiden Kodierenden. Besteht eine größere Abweichung, wurden die abweichenden Kodierungen diskutiert. Diese Diskussion erfolgte immer ergebnisoffen und es gab keine Pflicht, dass einer der Kodierenden seine Kodierung anpasst. Häufig wurde bei diesen Vergleichen Fehler in der Kodierung entdeckt und behoben. Mehrfach trat das Szenario auf, dass die beiden Kodierenden nicht übereingestimmt haben und bei ihrer jeweiligen Meinung blieben. Dieses Vorgehen ist bei der Inhaltsanalyse üblich (bspw. Johri und Yang 2017; Hamann 2002).

Zur Überprüfung der Übereinstimmung der beiden Kodierenden wird in dieser Arbeit Krippendorffs Alpha als Maß angewendet (Hayes und Krippendorff 2007; Krippendorff 2011). Dieses Verfahren ist robuster als prozentuale Übereinstimmung, da einzelne Abweichungen der Kodierenden weniger stark ins Gewicht fallen. Krippendorff empfiehlt für eine aussagekräftige Analyse eine Übereinstimmung von  $\alpha \geq .800$  und mindestens  $\alpha \geq .667$ , um vorsichtige Schlüsse ziehen zu können (Krippendorff 2004).

### **3.5 Aufbau des Softwaresystems als Grundlage für Studie 1**

Das für die Durchführung der Studie notwendige System wurde eigens für die Studie bzw. die Nutzung im Projektkontext EmployID erstellt. In diesem Abschnitt werden die dazugehörigen Komponenten beschrieben.

Das Gesamtsystem basiert aus zwei Einzelsystemen, die als Client- / Server-Struktur aufgebaut sind. Die Serverarchitektur ist in Abschnitt 3.5.2 und die Clientseite in Abschnitt 3.5.3 beschrieben. Insgesamt verfolgt der Ansatz die Idee keine neuen Systeme für Nutzer, die ohnehin viele Systeme im beruflichen Alltag nutzen, hinzuzufügen, sondern Reflexionsunterstützung als eine Art Plug-in in die aktuell genutzten Systeme zu integrieren (Prilla und Blunk 2015). Der Vorteil dieses Ansatzes ist, dass Nutzer die bisherigen Systeme bereits kennen und nicht die Nutzung eines neuen Systems erlernen müssen. Auf diese Weise könnten Widerstände gegenüber neuer Software umgangen werden. Durch Verknüpfung der Reflexionsunterstützung mit den übrigen Funktionen der bisherigen Systeme, können Nutzer dabei unterstützt werden, den für die kollaborative Reflexion notwendigen Kontext herzustellen. Dazu wird die

Reflexionsunterstützung als Plug-in in das bisherige System eingefügt. Für eine umfangreichere Darstellung der Vorteile dieses Ansatzes siehe Abschnitt 2.9.2.

### 3.5.1 Übergeordnete Architektur

Für die Umsetzung des Gesamtsystems wurden zwei Einzelsysteme entwickelt: Ein Backend, das auf Basis von Java-EE über eine Webserviceschnittstelle die Möglichkeit bietet, für bestimmte Kontexte Prompts anzufordern, oder Inhalte zu qualitativ zu analysieren (siehe Abschnitt 3.5.2). Diese Prompts sollen in der Community of Practice Plattform genutzt werden (siehe Abbildung 3-1). Das System basiert auf dem Content-Management-System (CMS) WordPress, mit den Plug-ins BuddyPress (für umfangreichere Funktionalitäten für ein persönliches Profil) und bbPress (Forum). Auf dem System der Community of Practice Plattform befindet sich ebenfalls das selbst entwickelte Plug-in zur Verwaltung der Prompts (siehe Abschnitt 3.5.3.1). Dieses analysiert die aktuelle Diskussionssituation und fordert dazu einen möglichst passenden Prompt vom Prompt-Service Backend an. Für eine Übersicht, welche Prompts für eine bestimmte Situation vorgesehen sind, siehe Abschnitt 4.3.2.

In den folgenden Abschnitten werden die beiden Systeme beschrieben. Der Prompt-Service befindet sich in Abschnitt 3.5.2 und die Community of Practice Plattform wird in Abschnitt 3.5.3 thematisiert.



Abbildung 3-1: Übersicht der Architektur

### 3.5.2 Prompt-Service Backend

Das Backend ist auf einer Java-EE Architektur aufgebaut. Während sich für eine einfache Webserviceschnittstelle eine andere Plattform, wie bspw. JavaScript angeboten hätte, waren die Anforderungen des Systems anfangs nicht final bekannt. Aus diesem Grund wurde eine bewährte und dem Entwickler bzw. Autor bekannte Umgebung für die

Erstellung des Systems ausgewählt. Der unterliegende Aufbau der Software hat auf den Aufbau oder die Durchführung der Studien keinen Einfluss.

Das System besteht aus zwei Schichten: die Datenbank-Schicht und die Webservice-Schnittstelle. Letztere beinhaltet die Annahme von Eingaben, die Logik zur Bereinigung von fehlerhaften Eingaben und weiteren Funktionen. Eine Schicht für eine grafische Benutzerschnittstelle ist nicht vorhanden. Als Datenbank kommt MySQL zum Einsatz. Als Webserver wird TomCat genutzt.

### 3.5.2.1 Komponenten

In diesem Abschnitt werden die unterschiedlichen Pakete des Systems kurz beschrieben.

- **Domain:** In diesem Paket befinden sich die Entitätsklassen, die einzelne Datenbank-Einträge widerspiegeln. Hier sind bspw. die Klassen für Nutzer, Log-Events oder Prompts gesammelt.
- **Exception:** Um aussagekräftige Fehlermeldungen zu ermöglichen bzw. intern auf bestimmte Fehler reagieren zu können, wurden unterschiedliche Fehlertypen als Klassen abgebildet. Diese sind in einem Paket gruppiert.
- **Listener:** In diesem Package sind zwei Listener enthalten, die auf das Ein- bzw. Ausloggen, und das dazugehörige Erstellen und Auflösen einer Session reagieren. Die Listener achten darauf, ob die entsprechenden Ereignisse geschehen und erstellen entsprechende Log-Einträge in der Datenbank.
- **Repository:** Hier befindet sich die Datenbank-Zugriffsschicht. In den unterschiedlichen Klassen existieren verschiedene Methoden, die Informationen aus der Datenbank einlesen, oder sie bearbeiten, hinzufügen oder löschen. Aufgrund des Spring JPA-Frameworks (siehe Tabelle 3-3) wurden keine eigenen Datenbankabfragen geschrieben, da der Compiler Methodennamen in Queries übersetzt (bspw. *findByUsernameIgnoreCase*).
- **Service:** Dies sind Klassen bzw. Methoden, die eine Art Zwischenschicht, zwischen der Webservice-Schnittstelle und der Datenbankschicht abbilden. Oft rufen die Methoden die entsprechenden Methoden im Repository-Package auf, aber teilweise sind zusätzliche Überprüfungen notwendig, die an dieser Stelle abgehandelt werden können. Rückblickend ist diese Zwischenschicht nicht notwendig. In diesem Package befinden Klassen, um einen Log-in über LDAP (Lightweight Directory Access Protocol) bzw. ActiveDirectory zu ermöglichen.

- Util: Sammlung verschiedener Hilfsklassen, bspw. für die Eingabevalidierung, oder für die Überprüfung, welcher Log-in-Mechanismus genutzt wird.
- Web: In diesem Package liegen die eigentlichen Klassen, welche die Webservice-Schnittstellen anbieten. Innerhalb der Methoden werden die Eingaben validiert und die benötigten Informationen zurückgegeben. Eine Übersicht der erstellten und von außen nutzbaren Schnittstellen findet sich in Abschnitt 3.5.2.3.

Zusätzlich liegen im Haupt-Package des Programms Konfigurationsdateien, wie bspw. für den Log-in über LDAP oder ActiveDirectory, oder für die Konfiguration der Datenbankverbindung.

### 3.5.2.2 Genutzte Frameworks

In Tabelle 3-3 sind die genutzten Frameworks und ihre Aufgabe im Backend zusammengefasst. Im Wesentlichen wurden die Frameworks verwendet, um bestimmte Aspekte der Programmierung, bspw. die Kommunikation mit der Datenbank oder die Erstellung von Log-Dateien, zu vereinfachen.

**Tabelle 3-3: Verwendete Frameworks<sup>11</sup> in der Software**

Framework	Aufgabe
<b>Spring</b>	<p>Spring ist eine Sammlung von unterschiedlichen Frameworks. <i>Spring Boot</i> ermöglicht Anwendungen, ohne aufwendige Konfiguration starten zu können. Die web application resources Datei (WAR), sprich die fertige Web-Anwendung, wird von Spring Boot an den TomCat-Webserver weitergereicht. Um die Kommunikation mit der Datenbank zu vereinfachen, wird <i>Spring JPA</i> genutzt, was in Verbindung mit Hibernate (siehe unten) genutzt wird.</p> <p>Zur Erstellung der Webservices, wird das <i>Spring Web Services</i> Framework verwendet. Um einen sicheren Benutzerkontext zu ermöglichen, und damit bekannt ist, welcher Benutzer welche Prompts abgerufen hat, wird das <i>Spring Security</i> Framework genutzt. Darüber ist bspw. ein Log-in über ein persönliches Passwort je Benutzer möglich. Da anfangs im Projekt ggf. eine Verbindung zu anderen Log-in-Providern diskutiert wurde, wurde ebenfalls eine Log-in-Möglichkeit über LDAP bzw. ActiveDirectory geschaffen, was über das gleichnamige <i>Spring LDAP</i> Framework erzielt wurde.</p> <p>Mehr unter: <a href="https://spring.io/">https://spring.io/</a></p>

<sup>11</sup> Für alle URLs letzter Zugriff am 12.08.2019



Framework	Aufgabe
<b>Hibernate</b>	Hibernate ist ein Framework für Object-Relational-Mapping, bei dem Java-Objekte als Datenbank-Einträge gespeichert werden können bzw. umgekehrt Datenbank-Einträge eingelesen und als Java-Objekte verwendet werden können. Mehr unter: <a href="http://hibernate.org/">http://hibernate.org/</a>
<b>Apache log4j</b>	Dies ist ein Framework zur vereinfachten Erstellung von Log-Dateien. Das Framework wird im Projekt dazu genutzt, je nach Log-Level, Fehler und Benutzeraktionen in eine Log-Datei auf dem Server zu speichern. Mehr unter: <a href="https://logging.apache.org/log4j">https://logging.apache.org/log4j</a>
<b>slf4j</b>	Um die Arbeit mit den Logs weiter zu vereinfachen, wurde das Framework slf4j genutzt. Dieses stellt eine vereinfachte Schnittstelle für Logs zur Verfügung und nutzt im Hintergrund log4j. Mehr unter: <a href="https://www.slf4j.org/">https://www.slf4j.org/</a>
<b>Maven</b>	Das Maven-Framework hilft den Build-Prozess von Java-Programmen zu vereinfachen. In diesem Projekt wurden die übrigen Frameworks über das Project-Object-Model (POM) von Maven verwaltet. Darüber lassen sich bspw. die Frameworks in einer bestimmten Version fixieren. Mehr unter: <a href="https://maven.apache.org/">https://maven.apache.org/</a>
<b>TomCat-Maven</b>	Mit diesem Plug-in kann das Java-Projekt über Maven auf einem internen TomCat-Webserver gestartet werden, ohne dass dazu ein fester Webserver installiert sein muss. Mehr unter: <a href="http://tomcat.apache.org/maven-plugin.html">http://tomcat.apache.org/maven-plugin.html</a>

### 3.5.2.3 Webservice-Schnittstelle

In Tabelle 3-4 sind die Webservice-Schnittstelle und die dazugehörigen Funktionsaufrufe beschrieben. Es existieren unterschiedliche Möglichkeiten der Selektion, um eine passende Liste von Prompts abrufen zu können. Das Zielsystem hat dann die Möglichkeit, sich aus der Liste der möglichen Prompt einen passenden Prompt nach bestimmten Kriterien oder zufällig auszuwählen und anzuzeigen.

**Tabelle 3-4: Übersicht welche Funktionen die Schnittstelle zum Abrufen von Prompts bereitstellt. Für jede Funktion ist auch der Pfad für den Aufruf angegeben, und um welche Art von Aufruf es sich handelt.**

Pfad und Funktion	Http-Aktion
/log	POST
Ermöglicht die Erstellung eines Log-Eintrags über den Webservice, um bei Bedarf gesonderte Log-Einträge erstellen zu können.	
/prompt	POST
Erstellung eines neuen Prompts über den Webservice.	

<b>Pfad und Funktion</b>	<b>Http-Aktion</b>
/prompt/	GET
Ruft eine Liste aller Prompts ab.	
/prompt/getCategoryList	GET
Ruft eine Liste aller Kategorien der Prompts ab. Über die Kategorie können weitere Metadaten an Prompts angefügt werden. Dieses Feld wurde in den Studien nicht genutzt.	
/prompt/getLanguageList	GET
Ruft eine Liste aller vorhandenen Sprachen ab.	
/prompt/getLocationList	GET
Ruft eine Liste aller Orte ab, für die Prompts vorhanden sind. Zu den Orten zählen die Erstellung neuer Themen, neuer Antworten oder die Startseite.	
/prompt/getTargetList	GET
Ruft eine Liste aller existierenden Rollen ab, für die Prompts vorrätig sind. Bspw. Autor, Lesender oder Antwortender.	
/prompt/category/{category}	GET
/prompt/location/{location}	
/prompt/target/{target}	
Ruft eine Liste von Prompts ab, die zu einer/einem bestimmten Category / Location / Target gehören. Optional kann die Sprache spezifiziert werden. Ohne spezifizierte Sprache wird eine Liste englischer Prompts zurückgegeben.	
/prompt/goal/{goal}	GET
Ruft eine Liste von Prompts ab, die zu einem bestimmten Ziel passen (siehe Abschnitt 3.2 im Konzept). Optional kann die Sprache spezifiziert werden. Ohne spezifizierte Sprache wird eine Liste englischer Prompts zurückgegeben.	
/prompt/goal/{goal}/location{location}	GET
/prompt/goal/{goal}/location{location}/target{target}	
/prompt/goal/{goal}/location{location}/target{target}/category{category}	
Ermöglicht das Abrufen einer Liste von Prompts nach mehreren Kriterien. Optional kann die Sprache spezifiziert werden. Ohne spezifizierte Sprache wird eine Liste englischer Prompts zurückgegeben.	

Neben den beschriebenen Schnittstellen existieren noch unterschiedliche Funktionalitäten zum Abrufen, Erstellen, Aktualisieren bzw. Löschen von Nutzern bzw. Nutzergruppen. Nutzer können ebenfalls in Nutzergruppen sortiert bzw. entfernt werden. Da diese Funktionalität in den Studien nicht genutzt wurde und diese Informationen in der Studie im WordPress-System abgebildet wurden, werden diese hier nicht näher beschrieben.

### 3.5.3 Community of Practice Plattform auf Basis von WordPress

Aufgrund der Anforderungen im Projekt EmployID wurde eine Community of Practice Plattform erstellt (siehe Abschnitt 4.2 bzw. Abbildung 2-1). Da im Forschungsprojekt unterschiedliche Anforderungen hinsichtlich der Plattform existierten, wurde ein System gewählt, dass mehr Funktionalitäten abbildet als eine einfache Foren-Software. Da aus Zeitgründen kein komplettes System neu erstellt werden sollte, wurde WordPress als Basissystem ausgewählt. Die Vorteile des Systems sind, dass es weit verbreitet ist und es zahlreiche Plug-ins gibt, um bestimmte Funktionalitäten nachzurüsten. Die benötigten Funktionalitäten wurden als Plug-ins implementiert, um für verschiedene Anwendungsfälle die Möglichkeit zu haben die Plattform unterschiedlich zu konfigurieren.

**EmployID – Demo of the Social Learning Platform**

Home News Groups Members Help 2 Language: ▼

## Welcome to the ZRSZ community

Here you can get an overview what happened in the community since your last visit

### Recent News

- Kick-Off Event for the Community Platform
- New development in Ljubljana: Specializing on certain industries
- Case Study: Local Employment Partnerships in UK
- Welcome to the learning platform

### Topic of the Month

**How to minimize the 'creaming effect'?**  
Help us all in finding good strategies against the 'creaming effect'. This effect describes that when referring people the qualified ones are getting referred over and over and the less qualified ...

**Latest reply (by Alice A)**  
I often try to include one or more persons who are on paper less qualified but who were quite motivated in the appointments here at ZRSZ. This can account for a lot and I even got Feedback from ...

[Join the discussion](#)

### Recent Topics in the Forums

- My super secret topic
- How to minimize the 'creaming effect'?
- Topic of the Month Archive
- Employers making requests for candidates not eligible: What to do?
- Examples for motivating clients?

### Most useful posts in the last 30 days

<b>Not willing to take job - motivation?</b> There was a job vacancy at company X. If offered it to a candidat...	2 Votes
<b>Reply to: not-willing-to-take-job-motivation</b> The same happened to me recently. I offered a vacancy for assista...	2 Votes
<b>Reply to: how-to-minimize-the-creaming-effect</b> I often try to emphasize the social skills someone has when refer...	2 Votes
<b>Reply to: not-willing-to-take-job-motivation</b> This also happened many times to me and my colleagues here in Z. ...	1 Votes
<b>How to minimize the 'creaming effect'?</b> Help us all in finding good strategies against the 'creaming ef...	1 Votes

### Tags

#foreigners #motivation

### Your thought of the day

What are your recent work related issues? Write down your experiences or try to help your colleagues solving their problems

Post your thoughts in the:  
→ [General forum](#)

[Edit](#)

**Oliver**  
[Log Out](#)

Search Forums

Search for:  [Search](#)

Currently Online

Abbildung 3-2: Prompts auf der Startseite der Plattform (unten, gelb hinterlegt)

WordPress<sup>12</sup> ist ein Content Management System (CMS), welches hauptsächlich für Blogs und Webseiten genutzt wird. Zur Erstellung einer Diskussionsumgebung wird das Plug-in bbPress<sup>13</sup> genutzt. Da Nutzer ein eigenes Profil auf der Seite pflegen können sollten, wurde ebenfalls das Plug-in BuddyPress<sup>14</sup> genutzt. Es wurden weitere existierende Plug-ins genutzt, um bspw. Funktionalitäten für E-Mails, Sicherheit, Caching, Mehrsprachigkeit, Logging, etc. abzubilden. Zusätzlich wurden im Rahmen des Projekts EmployID unterschiedliche Plug-ins entwickelt, um bspw. die Startseite der Community of Practice Plattform zu unterstützen, um Moderatoren wöchentliche Statistiken über die Aktivität der Nutzer zu schicken, etc. (EmployID 2018, 72–73).

Zur Verwendung der Plattform ist es für Nutzer erforderlich, sich mit einem Account einzuloggen. Auf diese Weise kann in Evaluationen verfolgt werden, wie sich Nutzer verhalten oder weiterentwickeln. Zusätzlich war es eine Anforderung innerhalb des Forschungsprojekts, dass Nutzer im Falle von Missbrauch der Plattform identifizierbar sind.

Zur Durchführung der vorliegenden Studie wurden zwei eigene Plug-ins entwickelt: ein Plug-in zur Anzeige von Prompts und ein Plug-in zur Verwaltung der nutzergenerierten Prompts. Beide Plug-ins werden im Folgenden beschrieben.

### **3.5.3.1 Prompting-Plug-in**

Das Ziel dieses Plug-ins ist es dem Nutzer zur Unterstützung von Reflexion Prompts anzuzeigen. Die Prompts werden dabei anhand der unterschiedlichen Ziele aus dem Konzept ausgewählt (siehe Abschnitt 3.2). Zur Bestimmung, welches Ziel in einer gegebenen Situation das Richtige ist, analysiert das Prompting-Plug-in die aktuelle Gesprächssituation, den Ort des Nutzers auf der Community of Practice Plattform und die Rolle des Nutzers in einer Diskussion. Auf dieser Basis zeigt das Prompting-Plug-in unterschiedliche Prompts an. Diese unterschiedlichen Orte der Anzeige und die Rollen, sowie die Speicherung von Daten werden im Folgenden beschrieben.

Dieses Plug-in wird in Studie 1 genutzt (siehe Kapitel 4) und zeigt basierend auf unterschiedlichen Kriterien die Prompts an.

---

<sup>12</sup> <https://wordpress.org/> (Letzter Zugriff am 12.08.2019)

<sup>13</sup> <https://bbpress.org/> (Letzter Zugriff am 12.08.2019)

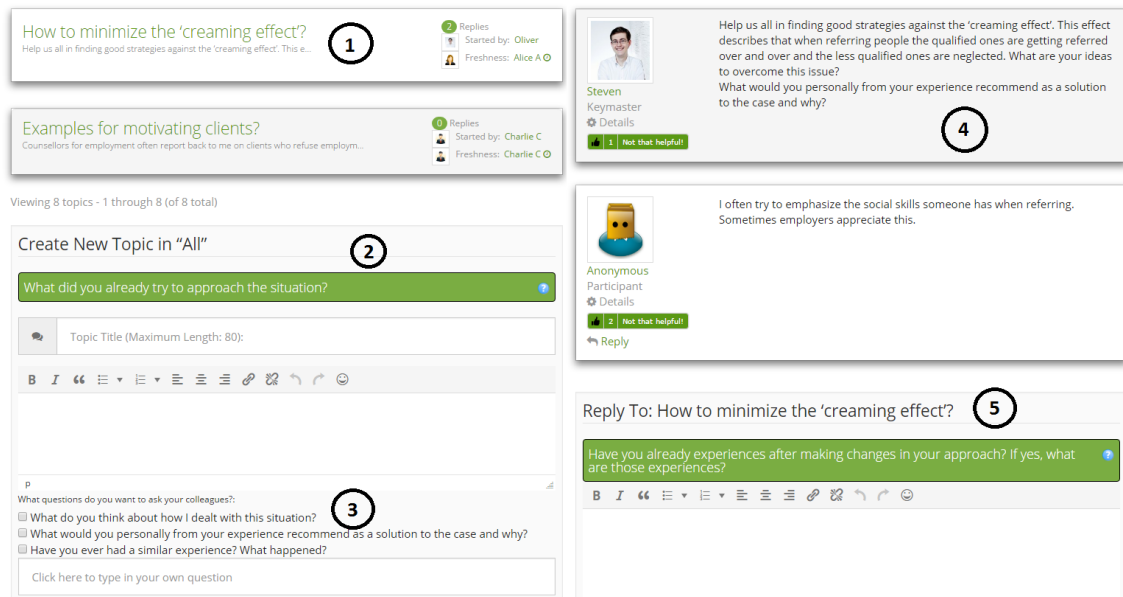
<sup>14</sup> <https://buddypress.org/> (Letzter Zugriff am 12.08.2019)

#### 3.5.3.1.1 Nutzung des Plug-ins in Studien

Zur Unterstützung von Studien ist es möglich, im Plug-in Zeiträume zu definieren, in denen Nutzer Prompts sehen bzw. nicht sehen können. Dazu werden Nutzer in unterschiedliche Gruppen einsortiert. Bei Aktivierung des Plug-ins innerhalb von WordPress werden zunächst alle existierenden Nutzer in eine Gruppe für die Studie einsortiert (siehe Abschnitt 4.3). Diese Einsortierung dient dazu alle Nutzer, die sich bereits vor dem Start der Studie auf der Plattform registriert haben mit in die Studie aufzunehmen. Die Einsortierung erfolgt pseudo-zufällig, sodass die Gruppengrößen annähernd vergleichbar bleiben, wobei jede Gruppe zwischen 75 %-125 % der Größe der anderen Gruppe sein darf. Dies verhindert Situationen, in denen mehrere Dutzend Personen zufällig in eine Gruppe sortiert und in der anderen Gruppe keine bis sehr wenige Personen vorhanden sind. Jede Person, die sich nach der Aktivierung des Plug-ins registriert, wird ebenfalls in eine der Gruppen einsortiert. Administratoren werden in eine dritte Gruppe sortiert, die immer (bei jedem Seitenaufruf) Prompts sehen kann, um die ordnungsgemäße Funktionsweise des Plug-ins kontrollieren zu können.

#### 3.5.3.1.2 Darstellung der Prompts

Das Ziel der Prompts ist die Unterstützung von Reflexion während des Austausches in der Community of Practice. Da der Austausch überwiegend über Beiträge in dem Forum stattfindet, sollen die Prompts sowohl das initiale Schreiben und Reflektieren beim Erstellen eines neuen Themas abdecken als auch die kollaborative Reflexion beim Erstellen von Antworten. Die Prompts auf der Startseite der Plattform sollen einerseits Reflexion anregen, aber auch Nutzer dazu anregen, neue Themen zu verfassen. Das Plug-in kann dazu Prompts an unterschiedlichen Orten anzeigen. Mit *Ort* ist in dieser Arbeit der Bereich auf der Webseite der Community of Practice Plattform gemeint. Zur Zeit werden drei Orte unterstützt: auf der Startseite (Abbildung 3-2), in der Liste der Themen (bei Erstellung eines neuen Themas; Abbildung 3-3 links) und innerhalb von Themen (bei der Erstellung von neuen Antworten; Abbildung 3-3 rechts).



**Abbildung 3-3: Prompts im Bereich der Foren (grün hinterlegt)**

In dem Forenbereich gibt es zwei Möglichkeiten der Darstellung: nahe der Liste der Diskussionsthemen und innerhalb einer Diskussion. Die Abbildung 3-3 zeigt auf der linken Seite die Prompts (2) unterhalb der Liste der Diskussionsthemen (1). Da beim Erstellen eines neuen Themas noch keine Diskussionsrollen existieren, gibt es dort keine Unterscheidung unterschiedlicher Rollen. Die Prompts sind nur sichtbar, wenn man aktuell laut Studiendesign Prompts sehen soll (für die Anwendung des Plug-ins in der Studie siehe Beschreibung der Studien in den Abschnitten 4.4.1 und 4.5.1). Auf der rechten Seite befinden sich die Prompts (5) unterhalb der Diskussion (4). Der Prompt ist dort platziert, wo der Nutzer einen Beitrag schreibt und damit die Möglichkeit hat, den Prompt während des Verfassens zu sehen. Der Autor hat die Möglichkeit während des Verfassens eines neuen Themas eine, keine oder mehrere Fragen für seine Kollegen auszuwählen (3) (Beschreibung des Plug-ins in Abschnitt 3.5.3.2). Diese Frage wird automatisch ans Ende des Themas angefügt und ist für alle sichtbar. Für die Nutzer, die laut Studiendesign aktuell in der Phase sind, in der Prompts sichtbar sind, wird der vom System ausgewählte Text des Prompts mit der vom Autor ausgewählten Frage überschrieben (nicht in der Abbildung visualisiert). Auf diese Weise kann der Autor eines Themas selbst Prompts an seine Kollegen schicken und so vermutlich die Diskussion besser steuern. Ein Beispiel für diesen Aspekt adaptierten Prompt ist der Prompt, der darauf abzielt Feedback einzuholen (Z7). Dieses Feature soll Reflexion unterstützen, indem der Autor gezielt Fragen stellt (bspw. Daudelin 1996). Diese Art von Prompts wird im Weiteren als *nutzergenerierte Prompts* bezeichnet.

Prompts auf der Startseite der Plattform sind in Abbildung 3-2 (unten in der Rubrik „Your thought of the day“, gelb hinterlegt) dargestellt. Das Ziel dieser Prompts ist, die Personen,

die sich auf der Startseite einen Überblick über die neuesten Themen und Diskussionen verschaffen, anzuregen einen neuen Beitrag zu verfassen. Zur Umsetzung des Prompts wurde auf der Startseite ein sogenanntes Widget platziert, welches den Prompt-Text beinhaltet. Dieses prüft beim Seitenaufruf, ob der Nutzer aktuell Prompts sehen darf. Dabei wird geprüft, ob der Nutzer laut Studie in der entsprechenden Experiment-Kondition ist, für die zum gegenwärtigen Zeitpunkt Prompts angezeigt wird. Falls dies der Fall ist, fordert das Plug-in vom Prompt-Service einen zufälligen Prompt für diesen Ort an. Für die Startseite gibt es keine detaillierte Analyse der Situation, wie es innerhalb der Themen durchgeführt wird. Zusätzlich erstellt das Prompting-Plug-in einen passenden Log-Eintrag für diese Situation. Unterhalb der Anzeige des Prompts zeigt das Plug-in einen Link zum Forenbereich an. Wenn man auf den Link unterhalb des Prompts klickt, dann wird man in den Forenbereich zur Erstellung eines neuen Themas weitergeleitet (siehe Abbildung 3-3 links) und der Prompt wird beibehalten und dort ebenfalls angezeigt. Das System speichert über einen URL-Parameter, dass der Nutzer von der Startseite kommend den Link genutzt hat. Auf diese Weise ist nachvollziehbar, falls ein Nutzer über den Link des Prompts auf der Startseite in den Forenbereich springt und dort einen Beitrag verfasst. Falls der Prompt bei dem Empfänger eine Idee anregt und dieser einen anderen Pfad in den Forenbereich nutzt, ist dies nicht anhand der Logdaten nachvollziehbar.

#### 3.5.3.1.3 Identifikation von Rollen

Das Softwaresystem ist in der Lage Prompts an Personen, die in einer Diskussion eine bestimmte Rolle einnehmen, zu schicken. Das Ziel dieser Mechanik ist einerseits, eine möglichst passende Hilfestellung für Reflexion zu bieten und andererseits, um zu vermeiden, dass Benutzer aufhören die Prompts zu beachten, nachdem sie für den jeweiligen Kontext stark unpassende Prompts bekommen haben.

Zu den Rollen gehört der *Autor* eines Diskussionsthemas. Die Annahme ist, dass derjenige, der das Thema initial erstellt, ein Ziel verfolgt. Auf Basis der Modelle zu Reflexion liegt nahe, dass der Autor nach Ideen und Erfahrungen sucht, um bspw. ein Problem zu lösen, oder eine zukünftige Handlung plant (bspw. Boud, Keogh und Walker 1985). Dementsprechend sieht der Autor im Gegensatz zu jemandem, der noch nicht geantwortet hat, Prompts, die ihn in die Richtung führen sollen, konkrete Schritte für die Zukunft zu benennen. Weiterhin gibt es die Rolle des *Lesers*, zu der Personen gezählt werden, welche die Diskussion gelesen aber in dieser noch nicht geantwortet haben. Hier ist ein Ziel die Personen zur Teilnahme zu bewegen. Die letzte Rolle ist die des *Antwortenden*: Jemand, der bereits eine Antwort verfasst hat. Der Autor behält die Rolle

des Autors, selbst wenn er in seinem eigenen Thema eine Antwort verfasst hat (bspw. in Reaktion auf den Beitrag eines Dritten).

#### 3.5.3.1.4 Technischer Ablauf und Speicherung der Ereignisse

In einer Log-Tabelle speichert das Plug-in, welcher Nutzer, zu welchem Zeitpunkt, in welchem Teil der Seite, welchen Prompt gesehen hat.

Beim Aufruf der Liste der Themen in einem Forenbereich prüft das Plug-in, ob der Nutzer, basierend auf der Zugehörigkeit zu der Gruppe in der Studie, zum aktuellen Zeitpunkt Prompts sehen soll. Falls dies der Fall ist, ruft es für diesen Ort einen neuen Prompt vom Prompt-Service ab. Anschließend wird ein passender Log-Eintrag erstellt. Dieser enthält zusätzlich einen Hash-Wert der aktuellen Session des Nutzers. Sollte der Nutzer sich dann entscheiden einen neuen Beitrag zu verfassen, wird auf Basis des Session-Hash-Wertes der vorherige Log-Eintrag aktualisiert.

Das Vorgehen für Prompts innerhalb von Themen, sprich beim Verfassen von Antworten, ist leicht anders. Hier wird ebenfalls zunächst ermittelt, ob der Nutzer Prompts sehen darf. Falls das Plug-in für nutzergenerierte Prompts aktiviert ist (siehe Abschnitt 3.5.3.2) und der Autor des Themas einen nutzergenerierten Prompt für das Thema ausgewählt hat, dann wird dieser Prompt angezeigt. Falls der Autor das nicht getan hat, wird ein passender Prompt für diese Situation ermittelt. Hierfür wird die Rolle des Nutzers ermittelt. Die Rolle gibt an, ob der aktuelle Nutzer der Autor des aktuellen Themas ist, ob er bisher nur gelesen hat, oder ob er bereits in diesem Thema eine Antwort verfasst hat. Je nach Rolle (siehe Abschnitt 4.3.2), ruft das Plug-in für diesen Ort einen Prompt vom Prompting-Server ab. Für diesen Vorgang wird ebenfalls ein Log-Eintrag erstellt und bei Bedarf, analog wie bei Prompts für neue Themen, aktualisiert.

Das Plug-in aktualisiert den Eintrag, sobald ein Lesezugriff zu einem Schreibzugriff wird, sprich der Nutzer ein Thema oder eine Antwort verfasst hat.

#### 3.5.3.1.5 Verwaltung und Verhalten im Fehlerfall

Falls der Prompt-Service nicht erreichbar ist, bspw. weil ein unerwarteter Fehler im System aufgetreten ist, wird ein Basis-Prompt angezeigt (siehe Abschnitt 4.3.2). Dieser Basis-Prompt wird am selben Ort und in der Darstellung angezeigt (siehe Abbildung 3-2 und Abbildung 3-3), wie die eigentlichen Prompts, um eine reibungslose Nutzung der Plattform zu ermöglichen.

Das Plug-in ist auf Mehrsprachigkeit ausgelegt und zeigt die Prompts auf Englisch oder auf Slowenisch bzw. Kroatisch (je nach Anwendungsfall) an, je nachdem welche Sprache



der Nutzer ausgewählt hat. Dazu wird vom Prompt-Service die entsprechende Sprache eines Prompts angefordert.

Über eine Benutzerschnittstelle im Administrationsmenü von WordPress lassen sich die Zeiträume für die Studien anpassen.

### **3.5.3.2 Plug-in für nutzergenerierte Prompts**

Dieses Plug-in zeigt bei jedem Seitenaufruf des Forenbereichs aller Nutzer bei der Erstellung von neuen Themen eine Auswahlmöglichkeit vordefinierter Fragen und ein Eingabefeld für eigene Fragen an (siehe Abbildung 4-2 (3)). Zum einen weist dieses Plug-in den Autor darauf hin, dass es hilfreich sein kann eine Frage zu stellen. Zum anderen kann der Autor mithilfe der Frage leichter die Diskussion steuern.

Wenn der Nutzer ein neues Thema verfasst, werden die ausgewählten vordefinierten Fragen oder der selbst eingegebene Text an das Ende des Themas angehängt. Sollte der Nutzer zusätzlich in der Gruppe sein, die zu dem Zeitpunkt Prompts sehen kann, so werden die vom Autor ausgewählten Fragen bzw. der eingegebene Text zusätzlich als Prompt angezeigt (und eventuelle vom System generierte Prompts werden nicht angezeigt; siehe 3.5.3.1).

Das Plug-in ist auf Mehrsprachigkeit ausgelegt und zeigt die Prompts sowohl auf Englisch oder auf Slowenisch bzw. Kroatisch (je nach Anwendungsfall) an, je nachdem welche Sprache der Nutzer ausgewählt hat. Über eine Benutzerschnittstelle im Administrationsbereich von WordPress lassen sich die Texte für die vordefinierten Prompts anpassen.

## **3.6 Crowdsourcing als Umgebung für Studie 2**

Da Crowdsourcing als Basis für die zweite Studie genutzt wird, wird an dieser Stelle ein Überblick über die Umgebung gegeben. Crowdsourcing ermöglicht die Verteilung einer Aufgabe auf viele Personen und erfreut sich großer Beliebtheit bei Forschern. In diesem Abschnitt wird der Begriff definiert, aktuelle Nutzungsmöglichkeiten von Crowdsourcing und die Vorteile und Nachteile aufgezeigt.

### **3.6.1 Einführung**

Die meiste Literatur führt den Begriff *Crowdsourcing* auf einen Artikel von Howe im Wired-Magazin zurück (Howe 2006). In seinem persönlichen Blog versucht dieser sich

an einer ersten Definition des Begriffs: Howe fasst Crowdsourcing als eine Form des Outsourcings zusammen, bei der Unternehmen Aufgaben an Externe vergeben, welche diese überwiegend alleine oder kollaborativ lösen (Howe 2006). Der Unterschied zu Outsourcing ist, dass es einen offenen Aufruf für die Aufgabe gibt und dass es ein großes Netzwerk an Personen gibt, die an der Aufgabe arbeiten (Howe 2016). Personen, die Aufgaben im Bereich von Crowdsourcing erledigen, werden oft Worker bzw. Crowdworker oder auch Turker (auf der Plattform Amazon Mechanical Turk) genannt (Martin u. a. 2014). Häufig besteht eine Aufgabe im Crowdsourcing aus sehr vielen kleinen Teilaufgaben und Crowdworker können davon beliebig viele durchführen.

Es lässt sich zwischen freiwilligen und bezahlten Tätigkeiten unterscheiden. Wikipedia<sup>15</sup> ist ein Beispiel freiwilliger Tätigkeit, bei dem viele Autoren unbezahlt Artikel schreiben oder versuchen diese zu verbessern (Wagner und Prasarnphanich 2007). Man kann die Forschungstätigkeit der NASA<sup>16</sup> unterstützen, indem man freiwillig Krater auf dem Mars markiert (Moon und Sproull 2008), oder biologisch/chemische Forschung beim Falten von Proteinen unterstützen (Savage 2012). Bei solchen Tätigkeiten ist die Teilnahme häufig intrinsisch motiviert (Wagner und Prasarnphanich 2007). Bei bezahlten Tätigkeiten ist das Bild nicht exakt eindeutig und manche Autoren verweisen darauf, dass die Personen die Aufgaben nicht aufgrund der monetären Kompensation erledigen (Buhrmester, Kwang und Gosling 2011), während hingegen andere Autoren darauf hinweisen, dass die Personen, die im Bereich Crowdsourcing tätig sind, in Foren häufig und ausführlich über faire Entlohnung diskutieren (Martin u. a. 2014). Andere Autoren führen Geld als Hauptgrund an (Gadiraju, Kawase und Dietze 2014). Es gibt es Personen, die auf das Einkommen aus Crowdsourcing-Arbeit angewiesen sind (Martin u. a. 2014) bzw. Crowdsourcing als Haupteinkommensquelle nutzen (Ross u. a. 2010). Geld ist die wichtigste Form der Entlohnung (Gadiraju u. a. 2015).

Vermeint wird Crowdsourcing in der Wissenschaft eingesetzt, da es für Forscher sehr einfach geworden ist, in kurzer Zeit sehr viele Probanden zu akquirieren (Behrend u. a. 2011; Cole u. a. 2009; Crump, McDonnell und Gureckis 2013; Heer und Bostock 2010; Mason und Watts 2009; Amir, Rand und Gal 2012; Cole u. a. 2009; Heilman und Smith 2010; Horton, Rand und Zeckhauser 2011). Die Qualität der Arbeitsergebnisse wird in Abschnitt 3.6.6 thematisiert. Durch die leichtere Akquise von Probanden sparen Forscher Zeit (Kittur, Chi und Suh 2008) bzw. sind in der Lage an den Kosten für Assistenten, Laboreinrichtung bzw. –Nutzung und Reiseentschädigung für Probanden zu sparen (Horton, Rand und Zeckhauser 2011). Einsparungspotenzial ist bei Probanden möglich,

---

<sup>15</sup> <https://www.wikipedia.org/> (Letzter Zugriff am 12.08.2019)

<sup>16</sup> <https://nasaclickworkers.com/classic/> (Letzter Zugriff am 29.05.2019)

da Forscher über Crowdsourcing leichter für die real genutzte Zeit zahlen können und Kosten teils um den Faktor zehn niedriger sind (Heer und Bostock 2010). Einzelne Studien haben sehr umfangreiche Aufgaben für sehr geringe Beträge (0.02 \$) durchführen lassen (Buhrmester, Kwang und Gosling 2011). Hier ist es wichtig, darauf zu achten, dass die Probanden für ihre Arbeit fair entlohnt werden. Dies ist ein häufiges Thema unter den Personen, die Crowdsourcing-Aufgaben erledigen (Martin u. a. 2014).

Häufige Aufgaben in Crowdsourcing sind Annotation von Bildern, Relevanzbewertung, Stimmungsanalyse in Texten, Übersetzungsarbeiten, Erkennen von Duplikaten, Transkription oder Objektklassifizierung (Difallah, Demartini und Cudré-Mauroux 2012; Martin u. a. 2014). Häufig kommt Crowdsourcing für Aufgaben infrage, die aufgrund großer Datenmengen sehr umfangreich sind und die von Computern (noch) nicht alleine durchgeführt werden können (Savage 2012; Difallah, Demartini und Cudré-Mauroux 2012; Gadiraju u. a. 2015). Das Annotieren großer Datenmengen ist häufig ein Vorbereitungsschritt, um Algorithmen für maschinelles Lernen zu trainieren (Fails und Olsen 2003). Aber nicht nur das Aufbereiten bzw. Klassifizieren von Daten gehört zum Aufgabenumfang, sondern auch das Erstellen von Inhalten, wie bspw. Generieren von neuen Ideen für Unternehmen, wird über Crowdsourcing in Auftrag gegeben (Blohm u. a. 2016). Es ist es in der Praxis unerheblich, welche Person, welchen Anteil in der Bearbeitung einer Aufgabe hat. Beispielsweise kann für die Aufgabe des Annotierens von Bildern eine Person nur eine Handvoll Bilder annotieren, während eine andere beteiligte Person einen Großteil der Bilder annotiert. Beide werden entsprechend ihrem Arbeitsumfang entlohnt.

Die unterschiedlichen Tätigkeiten, die Crowdworker erledigen, lassen sich wie folgt klassifizieren (Gadiraju u. a. 2015):

- *“Information Finding”*: Auffinden von benötigten Informationen, wie bspw. den zweiten Vornamen von Persönlichkeiten (Gadiraju, Fetahu und Kawase 2015).
- *“Verification and Validation”*: Verifizieren von Informationen, wie bspw. Spam-SMS aus einer Auswahl finden (Gadiraju, Fetahu und Kawase 2015).
- *“Interpretation and Analysis”*: Informationen basierend auf eigener Interpretation geben, wie bspw. Bewertung von Ideen (Blohm u. a. 2016).
- *“Content Creation”*: Erstellung von neuen Inhalten, wie bspw. Erstellung von Videos (Doan und Ramakrishnan 2014).
- *“Surveys”*: Umfragen, häufig in Studien oder für Marktanalysen genutzt, wie bspw. Studien zur Demografie und Persönlichkeit von Crowdworkern (Kazai, Kamps und Milic-Frayling 2012).
- *“Content Access”*: Aufruf von Online-Inhalten.

### 3.6.2 Einsatzbereiche von Crowdsourcing

Crowdsourcing wird nicht nur in einem professionellen Rahmen genutzt, sondern teilweise findet sich die Arbeitsumgebung in Spielen versteckt wieder, wie beispielsweise bei dem Google Image Labeler, der als *game with a purpose* bezeichnet werden kann (von Ahn und Dabbish 2008; Doan und Ramakrishnan 2014). Crowdsourcing wurde auch genutzt, um bei der Suche nach Vermissten Satellitenbilder auszuwerten (Fost 2008).

Im kommerziellen Bereich wird Crowdsourcing oft genutzt, um Kosten zu reduzieren bzw. um Expertise und Kreativität von Externen einzukaufen (Pedersen u. a. 2013), bspw. falls interne Abteilungen für Forschung und Entwicklung bestimmte Probleme nicht lösen können (Lakhani u. a. 2007).

Im wissenschaftlichen Bereich hilft Crowdsourcing, wie bereits beschrieben, um insbesondere leicht an eine große Menge an Probanden zu gelangen. Im Bereich der Psychologie wird häufig von *webbased*-Experimenten gesprochen (Crump, McDonnell und Gureckis 2013; Germine u. a. 2012). Dadurch, dass bei jeder Aufgabe sich die Gruppe der Crowdworker dynamisch neu bildet, ist die Demografie in den Studien unterschiedlich (Paolacci und Chandler 2014). Anhand der ID eines Crowdworkers kann man diesen von der Durchführung von Aufgaben blockieren, was für Studien den Vorteil hat, dass man einstellen kann, welche Nutzer(gruppe) an einer (Folge-)Studie (erneut) teilnehmen dürfen (Paolacci, Chandler und Ipeirotis 2010).

In der Wissenschaft kann Crowdsourcing genutzt werden, um Aufgaben an Externe zu verteilen. Beispielsweise ist die Erkennung von möglichen Proteinstrukturen eine stark visuelle Aufgabe, die nach kurzem Training von Laien durchgeführt werden kann. Die University of Washington konnte so ein Problem, was bisher über ein Jahrzehnt ungelöst war mit einer spielerischen Anwendung, die auf Prinzipien von Crowdsourcing basiert, innerhalb von drei Wochen lösen (Savage 2012). Ein anderes Beispiel ist die Nutzung von Crowdsourcing zum Auffinden von Hydranten, um Verbesserungsmöglichkeiten für Notfalleinsatzkräfte zu schaffen (Vivacqua und Borges 2012). Crowdsourcing wurde ebenfalls in Ökonomie-Studien (Amir, Rand und Gal 2012), Studien zur Kognition (Germine u. a. 2012) oder verhaltenspsychologischen Studien genutzt (Crump, McDonnell und Gureckis 2013).

### 3.6.3 Definition

Um den Begriff von Crowdsourcing weiter zu schärfen, wird an dieser Stelle eine Definition erarbeitet. Der Begriff wurde von Howe in einem Wired-Artikel verwendet

(Howe 2006) und später in einem Blogeintrag definiert (Howe 2016). Der Begriff beschreibt die Verteilung einer Aufgabe an Externe, die dann einzeln oder gemeinsam eine Lösung erarbeiten. Die Besonderheit ist, dass die Aufgabe offen für alle sichtbar ist und jeder direkt teilnehmen kann und es ein großes Netzwerk an möglichen Personen gibt, die die Aufgabe erledigen (Howe 2016). Neben dem Begriff des Crowdsourcing, wurden noch unterschiedliche andere Begriffe für diese Form der Arbeitsverteilung gefunden, die sich aber nicht durchgesetzt haben, wie bspw. „collective intelligence“, „smart mobs“, oder „mass collaboration“ (Doan und Ramakrishnan 2014).

In wissenschaftlichen Definitionen finden sich bestimmte Elemente von Crowdsourcing häufig wieder. Oft wird es als Weitergabe der Arbeit an andere beschrieben, um eigene Ressourcen zu schonen (Vivacqua und Borges 2012; Doan und Ramakrishnan 2014). Die Weitergabe der Arbeit erfolgt online (Behrend u. a. 2011) und die Arbeit wird an Externe weitergegeben, die unabhängig vom Auftraggeber sind (Behrend u. a. 2011; Kleemann, Voß und Rieder 2008). Diese Crowdworker sind oft Teil eines größeren Netzwerks von Personen (Doan und Ramakrishnan 2014). Die Crowdworker werden für das Erledigen einer bestimmten, häufig sehr kleinen (Difallah, Demartini und Cudré-Mauroux 2012), Aufgabe entlohnt (Behrend u. a. 2011; Vivacqua und Borges 2012). Die beteiligten Personen sind bei jeder Aufgabe anders (Pedersen u. a. 2013; Moon und Sproull 2008; Paolacci und Chandler 2014). Es wird die Arbeit an menschliche Crowdworker gegeben und nicht von Maschinen oder Computern ausgeführt (Doan und Ramakrishnan 2014; Difallah, Demartini und Cudré-Mauroux 2012).

Zusammengefasst beschreibt die Definition von Difallah, Demartini und Cudré-Mauroux die oben genannten Aspekte und dient als Grundlage für diese Arbeit:

*„Crowdsourcing is the process of indirectly employing anonymous people over the internet, often for a nominative amount of money, to complete concise tasks (called micro-tasks) that are typically too complex for today's computers but relatively simple for humans.“*  
(Difallah, Demartini und Cudré-Mauroux 2012)

Der Hintergrund ist, dass die durchzuführenden Aufgaben, bspw. das Annotieren von Bildern eine Aufgabe ist, die für Menschen sehr leicht auszuführen ist, aber für heutige Computer noch sehr schwierig durchzuführen ist. Dies liegt bspw. daran, dass gleiche Objekte aus unterschiedlichen Winkeln stark unterschiedlich aussehen. Dies erschwert die Erkennung für Computer. Gleichzeitig werden solche Aufgaben nicht von allen Unternehmen in großer Menge benötigt, wodurch es nicht notwendig ist Mitarbeiter für die jeweilige Aufgabe einzustellen.

Abzugrenzen ist der Begriff des Crowdsourcing von der Open Source Bewegung, bei der zwar viele Personen verteilt an einem Erzeugnis arbeiten, welches immer frei zugänglich ist, was bei Arbeitsergebnissen aus Crowdsourcing nicht der Fall ist (Brabham 2008; Vivacqua und Borges 2012). Im Gespräch mit Howe grenzt Larry Huston, Vizepräsident bei Procter & Gamble, den Begriff des Crowdsourcing von Outsourcing ab, da für ihn die Kollaboration mit den Personen nach Erledigung der initialen Aufgabe nicht endet (Howe 2006). Dies ist von der Art der Aufgabe abhängig, da Firmen wie Procter & Gamble eher auf die Plattform InnoCentive ausgerichtet sind, um dort klassische Forschungs- und Entwicklungsaufgaben durchführen zu lassen. In dem Falle ist eine Fortsetzung der Kooperation wahrscheinlicher als bspw., wenn viele Personen simple Bilder (bspw. Hunde und Katzen) für einen Algorithmus im Bereich maschinellen Lernens klassifizieren.

### 3.6.4 Plattformen für Crowdsourcing

Aktuell weit verbreitete Plattformen für Crowdsourcing sind insbesondere Amazon Mechanical Turk<sup>17</sup> (AMT), FigureEight<sup>18</sup> (ehemals Crowdfunder) oder InnoCentive<sup>19</sup>. AMT ist als erste Plattform mit dem Crowdsourcing-Prinzip verfügbar gewesen und ist aktuell Marktführer. Die Plattform InnoCentive wird eher zur Verteilung von Forschungs- und Entwicklungsaufgaben von großen Konzernen, wie bspw. „[...] Boeing, DuPont, and Procter & Gamble [...]“ (Howe 2006), genutzt. Firmen nutzen die Plattform, um Probleme in der eigenen Forschung und Entwicklung extern lösen zu können, und zahlen dafür eine höhere Entlohnung als auf anderen Plattformen üblich ist. Die Entlohnungen reichen von 10.000 \$ bis zu 100.000 \$ (Howe 2006; Kleemann, Voß und Rieder 2008; Lakhani u. a. 2007) oder in seltenen Fällen bis zu einer Million US-Dollar (Brabham 2008). Andere Plattformen wie bspw. iStockPhoto<sup>20</sup> sind ähnlich, mit dem Unterschied, dass dort ebenfalls nur spezialisierte Personen, in diesem Fall Fotografen, ihre Arbeit anbieten können (Howe 2006; Brabham 2008).

Eine Übersicht über mehrere Crowdsourcing-Plattformen und deren Funktionalitäten (wie bspw. Qualitätssicherung) und weiteren Informationen (bspw. Demografie, oder Qualifikation der Arbeiter) wurde von Vakharia und Lease erstellt (2013). Inhaltliche Gruppierungen von Crowdsourcing-Plattformen anhand von Kategorien, wie

---

<sup>17</sup> <https://www.mturk.com/> (Letzter Zugriff am 12.08.2019)

<sup>18</sup> <https://figure-eight.com/> (Letzter Zugriff am 12.08.2019)

<sup>19</sup> <https://www.innocentive.com/> (Letzter Zugriff am 12.08.2019)

<sup>20</sup> <https://www.istockphoto.com/de> (Letzter Zugriff am 12.08.2019)

individueller oder kollaborativer Arbeit, oder die Art der Nutzerrekrutierung findet sich in Doan und Ramakrishnan (2014).

### **3.6.5 Demografie der Crowdworker**

Um wissenschaftliche Studien in so einer Umgebung durchzuführen, ist es wichtig zu wissen, welche demografische Verteilung der Nutzergruppe vorliegt. Die meisten wissenschaftlichen demografischen Untersuchungen beziehen sich auf AMT. Mitarbeiter auf dieser Plattform können sich allerdings nur in den USA und in Indien Guthaben direkt auszahlen lassen. Mitarbeiter aus anderen Ländern müssen auf Amazon.com-Gutscheine zurückgreifen. Dementsprechend besteht die Demografie überwiegend aus Nutzern der beiden Länder (Paolacci, Chandler und Ipeirotis 2010). Es wurde in einer Umfrage auf der Plattform verzeichnet, dass bspw. der Anteil von Indern auf der Plattform zwischen 2008 und 2010 von 5 % auf 34 % stark zugenommen hat (Ross u. a. 2010).

Allgemein lässt sich festhalten, dass auf AMT, im Vergleich zu Internetpopulationen, überdurchschnittlich viele Frauen arbeiten (Mason und Suri 2011; Paolacci, Chandler und Ipeirotis 2010), aber bspw. dieser Anteil bei US-Crowdworkern höher ist als bei indischen Crowdworkern (Ross u. a. 2010).

Die Demografie auf Crowdsourcing-Plattformen stimmt nicht mit einzelnen Staaten überein (Suri und Watts 2011). Allerdings sind Crowdworker als Nutzergruppe diverser hinsichtlich demografischer Daten als reguläre Probandengruppen aus dem Internet (Paolacci und Chandler 2014) und diverser ist als typische Nutzergruppen an US Colleges, die gerne für Studien herangezogen werden (Buhrmester, Kwang und Gosling 2011). Gegenüber den Schülern/Studierenden haben Crowdworker bspw. häufig eine höhere Arbeitserfahrung (Behrend u. a. 2011). Diese Gruppen sind nicht in allen Aspekten vergleichbar und Ergebnisse sind nicht für die jeweils andere Gruppe generalisierbar (Ward 1993). Für Studien ist anzumerken, dass die interne Konsistenz der Antworten der Nutzergruppe in demografischen Umfragen sehr hoch ist (Mason und Suri 2011).

### **3.6.6 Qualität der Ergebnisse**

Obwohl die Aussicht auf eine schnellere Akquise einer sehr diverseren Probandengruppe bei gleichzeitiger Kostenreduktion verlockend ist, stellt sich die Frage, inwieweit die Ergebnisse aus Crowdsourcing-Jobs qualitativ hochwertig sind und ob diese mit Ergebnissen aus Laborstudien vergleichbar sind.

In unterschiedlichen Bereichen wurde festgestellt, dass Ergebnisse aus Crowdsourcing-Jobs ähnlich gute Ergebnisse liefern wie Nutzer aus dem Labor: Evaluation von Nutzerinterfaces (Komarov, Reinecke und Gajos 2013), Bewertung von computer-generierten Fragen (Heilman und Smith 2010), Entscheidungsfindung (Paolacci, Chandler und Ipeirotis 2010), Wahrnehmung von prozentualen Größenunterschieden in Grafiken (Heer und Bostock 2010), Vervollständigung von Sätzen in linguistischen Studien (Schnoebelen und Kuperman 2010), Ökonomie-Experimente (Amir, Rand und Gal 2012), oder Annotation von Bildern (Nowak und Rüger 2010). Andere Autoren, die mehrere unterschiedliche Studien durchgeführt haben, berichten ebenfalls über die ähnlich hohe Qualität von Crowdsourcing-Ergebnissen (Gosling u. a. 2004; Horton, Rand und Zeckhauser 2011; Buhrmester, Kwang und Gosling 2011). In psychologischen Studien, wie bspw. zur Kognition (Germine u. a. 2012) oder zum Verhalten (Crump, McDonnell und Gureckis 2013), wird über eine hohe Qualität der Ergebnisse berichtet.

Allerdings zeigt sich, dass die Qualität nicht in allen Studien und bei allen Demografien gleichbleibend hoch ist. Asiatische Crowdworker liefern schlechtere Arbeitsergebnisse ab (Kazai, Kamps und Milic-Frayling 2012). In einer Studie wird berichtet, dass Crowdworker Schwierigkeiten mit sehr komplexen Aufgaben und mit Verbesserungen von Grammatik bzw. Korrekturlesen haben, wodurch in der Studie bis zu 30 % der Ergebnisse nicht nutzbar für Endverbraucher waren (Bernstein u. a. 2015). Die Qualität schwankt je nach Studie und es müssen zwischen zwei Prozent und fünfzig Prozent der Ergebnisse revidiert werden (Komarov, Reinecke und Gajos 2013; Heer und Bostock 2010; Kittur, Chi und Suh 2008). Es bieten sich unterschiedliche Mechanismen an, um Nutzer mit Täuschungsabsicht (siehe Abschnitt 3.6.7) zu erkennen bzw. Crowdworkern das Erledigen der Tätigkeit zu erleichtern (siehe Abschnitt 3.6.9).

Festzustellen ist auch, dass eine höhere Bezahlung nicht direkt mit einer höheren Qualität der Ergebnisse einhergeht (Buhrmester, Kwang und Gosling 2011), obwohl die höhere Bezahlung dazu führt, dass der Job schneller erledigt wird (Mason und Watts 2009), da vermutlich mehr Crowdworker interessiert an dem Job sind. Es ist Aufgabe des Auftraggebers, die Qualität zu sichern, was aufgrund großer Datenmengen manuell nicht bei jedem Job zu leisten ist (Difallah, Demartini und Cudré-Mauroux 2012).

### **3.6.7 Umgang mit Nutzern mit Täuschungsabsicht**

Ein häufiges Problem bei Crowdsourcing ist die Täuschungsabsicht mancher Nutzer: Um Aufgaben schneller abzuschließen, um somit mehr Geld zu verdienen, versuchen sie durch Eingabe von beliebigen Werten die Aufgabe abzuschließen (siehe *fast-deceiver* in



(Gadiraju u. a. 2015)). Im Folgenden werden einige Maßnahmen vorgestellt, um Täuschung zu verhindern bzw. einzudämmen.

Die einfachste Möglichkeit, Täuschung zu verhindern, ist die betreffenden Nutzer nicht an dem Job teilnehmen zu lassen bzw. vorher zu blockieren (Doan und Ramakrishnan 2014). Das ist auf den gängigen Plattformen möglich, allerdings muss dazu der entsprechende Nutzer erst identifiziert werden. Auf den gängigen Plattformen ist es möglich, Aufgaben nur für Nutzer eines bestimmten Ranges anzubieten, welcher sich aus vorherigen Aufgaben ergibt (Doan und Ramakrishnan 2014; Difallah, Demartini und Cudré-Mauroux 2012).

Nachdem eine Aufgabe erledigt wurde und der Auftraggeber die Antworten von niedriger Qualität identifiziert hat, kann post-hoc eine Liste mit blockierten Nutzern erstellt werden (Paolacci, Chandler und Ipeirotis 2010; Doan und Ramakrishnan 2014). Ein Ansatz ist, einen Testkatalog von Fragen vor das eigentliche Erledigen des Jobs zu schalten, um die Nutzer zu filtern, die eine Täuschungsabsicht haben. Dies wird oft als *Gold-Standard-Frage* bezeichnet und meint Fragen, auf die eine Antwort vorher bekannt ist (Difallah, Demartini und Cudré-Mauroux 2012; Oleson u. a. 2011; Doan und Ramakrishnan 2014). Die Herausforderung bei diesen Testfragen ist, dass sie nicht zu schwer sein dürfen, damit echte Crowdworker sie noch beantworten können und nicht zu leicht sein dürfen, damit sie nicht von Bots beantwortet werden können (Difallah, Demartini und Cudré-Mauroux 2012). Es werden viele Fragen in einem Testkatalog benötigt, damit Nutzer, die sehr oft einen bestimmten Job erledigen, diese nicht erkennen können (Difallah, Demartini und Cudré-Mauroux 2012). Zusätzlich passen sie nicht bei jeder Art von Job und es ist zu befürchten, dass Nutzer mit Täuschungsabsicht aufrüsten und diese Fragen in Zukunft leichter beantwortet werden können (Gadiraju u. a. 2015). Wenn man Aufgaben zu einem späteren Zeitpunkt wiederholt, kann man für die späteren Durchläufe Testfragen aus den bisherigen, korrekten Antworten generieren (Oleson u. a. 2011). Oleson empfiehlt das gezielte Einbauen von Fehlern, die beim Bearbeiten der Aufgabe aufgelöst werden, damit man den Trainingsprozess des Crowdworkers besser begleiten kann (Oleson u. a. 2011).

Manche Jobergebnisse, wie Antworten in Umfragen sind schwer überprüfbar und diese Aufgabentypen werden oft von Spammern erledigt (Komarov, Reinecke und Gajos 2013).

Es bietet sich die Nutzung von Überprüfungsfunktionen für Eingaben an, mit der Crowdworker direkt darauf hingewiesen werden, dass eine Eingabe fehlerhaft ist (bspw. Warnung bei Eingabe von Text in ein Zahlenfeld). Auf diese Weise können Personen, die beabsichtigen die Regeln der Aufgabe zu brechen (*rule-breaker*) bzw. schnell täuschen

wollen (*fast-deceiver*), daran gehindert werden allzu schnell an ihr Ziel zu kommen und ggf. häufigeres Erledigen derselben Aufgabe zu unterlassen (Gadiraju u. a. 2015). Damit werden fehlerhafte Ergebnisse zwar nicht ausgeschlossen, aber die Menge wird reduziert.

Ein weiterer Ansatz ist das Filtern der Ergebnisse. Bei Bewertungsaufgaben empfiehlt es sich, mehrere Bewertungen zu sammeln und dann diejenigen zu behalten, bei der mehrere Nutzer übereinstimmen (Oleson u. a. 2011). Falls Nutzer Aufgaben unnatürlich schnell erledigen, ist das ein Anzeichen für Täuschung und kann nachträglich leicht erkannt werden (Kittur, Chi und Suh 2008). Manche Plattformen bieten die Einstellungsmöglichkeit Antworten, deren Erstellungszeit unterhalb einer bestimmten Zeitdauer liegt, direkt zu verwerfen.

Letztlich lassen sich Nutzer mit Täuschungsabsicht nicht vermeiden und daher ist es wichtig, dass die Plattformen Daten nicht nur aggregiert, sondern auch als Rohdaten zur Verfügung stellt, damit man dort die entsprechenden als falsch identifizierten Antworten entfernen kann (Doan und Ramakrishnan 2014).

### **3.6.8 Nachteile von Crowdsourcing**

Neben den bereits beschriebenen Vorteilen von Crowdsourcing gibt es Aspekte, die den Nutzen bzw. das Prinzip Crowdsourcing einschränken.

Obwohl die Arbeitsergebnisse auf einem sehr hohen Level sind und Ergebnisse mit den Ergebnissen aus wissenschaftlichen Laborstudien vergleichbar sind (siehe Abschnitt 3.6.6), bleibt festzustellen, dass nicht jeder Crowdsourcerer für alle Aufgaben qualifiziert ist (Paolacci und Chandler 2014). Ebenfalls müssen die Nutzer mit Täuschungsabsicht bzw. deren Arbeitsergebnisse aus den gewünschten Daten entfernt werden (siehe Abschnitt 3.6.7). Da Crowdsourcing nur online durchgeführt wird, können nur Personen, die einen Computer besitzen bzw. damit umgehen können, an Crowdsourcing teilnehmen, was für Personen mit sehr niedrigem Einkommen nicht immer gegeben ist (Brabham 2009). Zusätzlich verfügt nicht jeder Nutzer über die gleiche Hardware, was einen Einfluss auf Studien nehmen kann (Komarov, Reinecke und Gajos 2013; Heer und Bostock 2010).

Crowdsourcing ermöglicht es, für geringe Beträge Arbeitsergebnisse zu erzielen, was das Geschäftsmodell von anderen qualifizierten Arbeitskräften angreift und Änderungen in deren Arbeitsumstände nach sich zieht. Ein Beispiel dafür ist das Aufkommen von Stock-Fotografie-Plattformen für Fotografen (Howe 2006; Brabham 2008).

### 3.6.9 Anforderungen an Studiendesign

Aufgrund der zahlreichen Studien, die im Bereich Crowdsourcing durchgeführt wurden, gibt es bereits unterschiedlichste Hinweise und Richtlinien für eine saubere Gestaltung von Crowdsourcing-Aufgaben. Zur Vermeidung von Täuschungsversuchen wurden unterschiedliche Maßnahmen bereits beschrieben (siehe 3.6.7).

Bei Design von Stimulus und Response-Aufgaben muss beachtet werden, dass die Hardware der Crowdworker unterschiedlich ist, und dass manche Crowdworker daher die Jobs nicht wie geplant durchführen können bzw. Zeiten zwischen manchen Durchführungen nicht vergleichbar sind (Crump, McDonnell und Gureckis 2013). Hier bietet es sich an den Nutzer zu fragen, ob es technische Schwierigkeiten bei der Durchführung der Aufgabe gab, um letztlich die Datenqualität zu erhöhen (Komarov, Reinecke und Gajos 2013). Der Crowdworker sollte in dem Fall trotzdem bezahlt werden (Komarov, Reinecke und Gajos 2013).

Ein großer Unterschied zwischen der Durchführung von Studien im Labor und über Crowdsourcing ist das Problem, dass Forschende in Crowdsourcing-Umgebungen keinerlei Kontrolle über die Umgebung haben und schlicht nicht wissen, ob bspw. der Crowdworker durch seinen TV abgelenkt ist (Crump, McDonnell und Gureckis 2013; Komarov, Reinecke und Gajos 2013; Gosling u. a. 2004). Typische Informationen zur Bildschirmnutzung, wie Bildschirm, Ausleuchtung, etc. sind selten überprüfbar (Heer und Bostock 2010).

Gleichzeitig ist es sehr schwierig, in einer Crowdsourcing-Umgebung dafür zu sorgen, dass, falls es die Studie so erfordert, die Probanden keinen Kontakt untereinander haben. In Laborexperimenten lässt sich dies kontrollieren, in dem man alle Probanden gleichzeitig einlädt und dadurch sicherstellt, dass keine Kommunikation stattfindet. Im Bereich von Crowdsourcing bleibt Forschern die Möglichkeit, die üblichen Crowdsourcing-Foren, wie bspw. Turkerc Nation (Martin u. a. 2014), zu beobachten und die Autoren von Diskussionen zu den Jobs der Forscher darum bittet, die Einträge zu depublizieren (Horton, Rand und Zeckhauser 2011). Falls die Aufgabe sehr schnell durch Crowdsourcing durchgeführt wurde, minimiert man die Zeit in der sich Probanden potenziell darüber austauschen können.

Wichtig für die Durchführung von Studien ist eine sehr präzise Aufgabenbeschreibung, damit Crowdworker exakt wissen, was von ihnen verlangt wird (Kittur, Chi und Suh 2008; Komarov, Reinecke und Gajos 2013; Martin u. a. 2014; Egelman, Chi und Dow 2014). Falls die Aufgaben sehr herausfordernd sind, hilft es Feedbackmechanismen zu implementieren (Crump, McDonnell und Gureckis 2013). Damit Crowdworker

längerfristig an einem Job Spaß haben, dürfen die Jobs nicht zu monoton gestaltet sein (Crump, McDonnell und Gureckis 2013). Die Drop-out-Rate, also die Rate in der Personen die Jobs abbrechen bzw. den Job nicht erneut anfangen, hängt von mehreren Faktoren, wie Länge der Aufgabe, finanzieller Entlohnung oder der persönlichen Motivation des Crowdworkers ab (Crump, McDonnell und Gureckis 2013).

Beim Vermitteln von Instruktionen wurde festgestellt, dass Crowdworker generell aufmerksamer sind als Nutzer im Labor (Hauser und Schwarz 2016), wobei nicht alle Crowdworker die Instruktionen vollständig lesen (Crump, McDonnell und Gureckis 2013). Crowdworker lernen bei manchen Jobtypen langsamer (Crump, McDonnell und Gureckis 2013). Um zu überprüfen, ob die Instruktionen gelesen und verstanden wurden bieten sich sogenannte *Instructional Manipulation Checks* an (Oppenheimer, Meyvis und Davidenko 2009), bei denen im Bereich der Instruktion dem Probanden angezeigt wird, dass man unten stehende Anweisung direkt ignorieren und anders reagieren solle als an der Stelle beschrieben steht. Damit lässt sich direkt überprüfen, ob die Personen die Instruktionen gelesen haben oder ob sie diese übersprungen und direkt mit der Bearbeitung eines Jobs begonnen haben.

## **4 Studie 1: Prompts als Reflexionsunterstützung in der Praxis öffentlicher Verwaltungen**

Diese Studie im Kontext öffentlicher Verwaltungen zielt darauf ab, das zuvor entwickelte Konzept (siehe Kapitel 3) in der Praxis zu erproben. Ein Teil dieser Studie wurde bereits veröffentlicht (Blunk und Prilla 2017b). Es wird das gleiche Set an Prompts in zwei unterschiedlichen Anwendungsgefällen getestet. Da sich beide Anwendungsfälle unterschiedlich entwickelt haben, werden die Datensätze zunächst getrennt analysiert und abschließend gemeinsam diskutiert. Die Grundannahme der Studie ist, dass Prompts eine Auswirkung auf die Antworten der Probanden haben können (siehe Abschnitt 2.10).

### **4.1 Kontext der Studie**

Das zuvor in Kapitel 3 entwickelte Konzept soll getestet werden. Es sollen textuelle Prompts zum Einsatz kommen, die unterschiedliche Aspekte kollaborativer Reflexion, wie bspw. das Referenzieren eigener Erfahrungen oder das Entwickeln neuer Lösungsvorschläge auslösen sollen. Die Studie läuft nicht unter Laborbedingungen oder in einer kontrollierten Umgebung ab, sondern wird in der betrieblichen Praxis von öffentlichen Verwaltungen durchgeführt.

Viele Studien setzen Prompts in einem kontrollierten Rahmen bzw. in definierten Aufgaben im Bildungsbereich ein (bspw. Renner u. a. 2016; Bannert 2006; Thillmann u. a. 2009; Davis und Linn 2000; Davis 2003). Bei solchen Studien könne äußere Einflüsse ein Stück weit kontrolliert werden. Die Teilnehmer bekommen Aufgaben, die mit Prompts versehen sind, wodurch garantiert ist, dass diese die Prompts sehen bzw. zu einer Interaktion gezwungen sind. Insbesondere im Bildungsbereich sind solche Studien in der Regel mit schriftlichen Berichten verbunden. In der betrieblichen Praxis ist die Teilnahme an einer Community of Practice oft freiwillig und daher existiert für die Probanden kein zwingender Grund, einen Text zu verfassen. Hinzu kommt, dass der Großteil der Mitglieder in der Peripherie einer Community angesiedelt ist (Wenger, McDermott und Snyder 2002, 56) und daher größtenteils nur lesend aktiv ist. Zusätzlich beinhalten die meisten Aufgaben im beruflichen Kontext keine Reflexion (Prilla u. a. 2015). Aus diesem Grund ist Reflexion in der Praxis ein zweitrangiger Prozess, der hinter der eigentlichen Aufgabenerfüllung zurücksteht. Zusätzlich existiert oft ein hoher Zeitdruck, der einen systematischen Ansatz für Reflexion behindern kann (Prilla u. a. 2015). Daher ist unklar, inwieweit sich Prompts zur Unterstützung kollaborativer Reflexion bzw. einzelner Aspekte kollaborativer Reflexion in der Praxis eignen. Ebenfalls ist unklar, ob diese eine ähnliche Wirkung haben, wie in Studien im Bildungsbereich gezeigt wurde.

Die Zusammenarbeit mit öffentlichen Verwaltungen fand im Rahmen des von der Europäischen Kommission geförderten Forschungsprojekts EmployID<sup>21</sup> statt. Im Rahmen des Projekts wurde das Konzept bei zwei Projektpartnern, den Arbeitsagenturen in Slowenien (siehe Abschnitt 4.4) und Kroatien (siehe Abschnitt 4.5), über jeweils mehrere Monate im Praxiseinsatz getestet. Eine Übersicht des Projektvorhabens befindet sich in Abschnitt 4.2.1. Die öffentlichen Verwaltungen haben im Rahmen des Forschungsprojekts Communities of Practice (siehe Abschnitt 2.1) zur Unterstützung kollaborativer Reflexion eingesetzt. Eine Beschreibung öffentlicher Verwaltungen als Kontext für die Entstehung einer Communities of Practice wurde bereits publiziert (Blunk und Prilla 2017a).

Die Forschungsfrage für diese Studie lautet, *inwieweit Prompts kollaborative Reflexion in der betrieblichen Praxis im Rahmen des Austausches in einer Community of Practice unterstützen können*. Die Studie zielt darauf ab, kollaborative Reflexion durch Prompts in der täglichen Praxis im Austausch in Communities of Practice einzusetzen. Betrachtet werden die Effekte von Prompts auf Personen, die neue Themen in eine Diskussion einbringen und auf diejenigen, die sich in der Diskussion beteiligen.

## **4.2 Projektkontext EmployID**

Da die Studie im Kontext des Forschungsprojekts EmployID durchgeführt wurde, wird an dieser Stelle zunächst die Zielstellung des Projekts und der Kontext der Studie selbst fokussiert.

### **4.2.1 Hintergrund und Zielstellung des Forschungsprojekts**

Die zunehmende Arbeitslosigkeit, der demografischen Wandel, die zunehmende Dynamik auf den Arbeitsmärkten und die fortschreitende Digitalisierung üben hohen Druck auf die Arbeitsagenturen und ihre Mitarbeiter aus (vgl. im Folgenden Bimrose u. a. 2014b; und Bimrose u. a. 2014a). In vielen Ländern ist eine häufige Änderung der Gesetzgebung ein Faktor, der Druck auf Arbeitsagenturen ausübt (Prilla u. a. 2014). Dieser Druck bedingt, dass die Mitarbeiter ihre Beratung an sich ändernde Gegebenheiten anpassen, wodurch sich ihre berufliche Identität wandelt (im Orig. *professional identity*). Die Zielsetzung des Projekts ist es, diesen beruflichen Wandel auf mehreren Ebenen bzw. über mehrere Ansätze zu unterstützen.

---

<sup>21</sup> Gefördert von der Europäischen Kommission in FP 7 (Projektnummer 619619).

Das Konzept der beruflichen Identität, dargestellt in Abbildung 4-1, besteht aus unterschiedlichen zusammenhängenden Elementen (Bimrose u. a. 2014b; Bimrose u. a. 2014a; Brown 1997). Es gibt in der beruflichen Identität sowohl eine persönliche als auch eine kollektive Ebene (Vgl. im Folgenden Bimrose u. a. 2014b; und Brown 1997). Die berufliche Identität umfasst berufliche Aktivitäten, den eigenen Bezug zum Beruf, zur beruflichen Kultur und den Weg, um diese Identität zu erreichen. Diese berufliche Identität ändert sich, sobald die Tätigkeit im Beruf einem Wandel unterliegt, weil bspw. neue Fertigkeiten auf individueller oder kollektiver Ebene benötigt werden. Diese Anpassung erfordert Lernen und oft gesonderte Unterstützung.

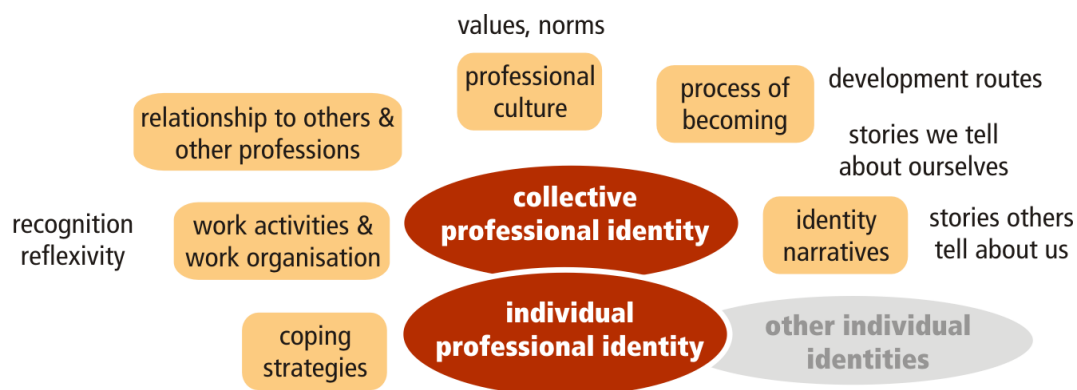


Abbildung 4-1: Theoriehintergrund in EmployID (Bimrose u. a. 2014a)

Eine der Möglichkeiten diesem Wandel in der beruflichen Identität zu begegnen ist (kollaborative) Reflexion (siehe Kapitel 2) (Prilla u. a. 2014). Diesen Prozess nennt man *career adaptability*, sprich die Fähigkeit, die eigene Karriere anzupassen (Prilla u. a. 2014). Reflexion ist eine passende Unterstützung für diesen Anpassungsprozess, da es einem erlaubt sich kritisch mit der eigenen Tätigkeit entweder in-situ, als *reflection-in-action* oder post-hoc in der Form von *reflection-on-action* auseinanderzusetzen (Schön 1983). Reflexion ist eine Möglichkeit durch informelles Lernen das Lernen am Arbeitsplatz zu unterstützen (Eraut 2004).

Es gibt vier unterschiedliche Aspekte der career adaptability: *challenging work*, *updating a substantive knowledge base*, *being self-directed and self-reflexive*, und *learning through interactions at work* (Brown u. a. 2012). Diese lassen sich alle durch Reflexion gezielt unterstützen (vgl. im Folgenden Prilla u. a. 2014): Bei der Bearbeitung herausfordernder Aufgaben kommen unweigerlich Situationen, in denen ad-hoc kein Lösungsansatz vorhanden ist oder wo sich die aktuelle Situation nicht so entwickelt hat, wie erwartet, was Reflexion auslöst (Boud, Keogh und Walker 1985). Das eigene Wissen aktuell halten wird zwar oft in formalen Trainings abgedeckt, ist aber ebenfalls über informelles Lernen in Form von Reflexion möglich (Eraut 2004). Selbstgesteuertes Lernen zielt eher auf individuelle statt kollaborative Reflexion ab. Hier ist es möglich durch Unterstützung

durch Software, bspw. zu Themen, welche die eigene berufliche Identität betreffen und die einer Reflexion bedürfen, zu sammeln und regelmäßig anzugehen. Der letzte Punkt, lernen durch Interaktionen am Arbeitsplatz, kann durch kollaborative Reflexion abgedeckt werden (siehe Abschnitt 2.5), da gezielt voneinander gelernt werden kann und sich Kollegen gegenseitig unterstützen können. Bei den oben beschriebenen Herausforderungen unterliegt das gesamte Team diesen Herausforderungen und kann sich diesen gemeinsam stellen.

#### **4.2.2 Verankerung der Studie im Projekt**

Insgesamt gab es eine Zusammenarbeit mit zwei unterschiedlichen Partnern im Projekt. In diesem Abschnitt wird die Hinführung auf den Einsatz von Community of Practice Plattformen im Projekt beschrieben. Es wird zunächst die Arbeitsagentur in Slowenien in den Fokus gerückt, weil dies der erste Partner war und in diesem Anwendungsfall die Community of Practice Plattform gemeinsam mit Nutzern gestaltet wurde. Bei den anderen Partnern wurde die dann bestehende Lösung für gleichartige Situationen ebenfalls bereitgestellt. Beide Arbeitsagenturen haben im Projekt mit anderen Forschungspartnern zusammengearbeitet. Eine konkrete Beschreibung des Umfelds bei den jeweiligen Partnern findet in den Abschnitten zum jeweiligen Anwendungsfall statt: Slowenien (siehe Abschnitt 4.4) und Kroatien (siehe Abschnitt 4.5).

Zu Beginn des Projekts wurden mit dem Projektpartner der slowenischen Arbeitsagentur (ZRSZ) zwei Workshops durchgeführt (vgl. im Folgenden Brown u. a. 2015, 17–21, 25–29). Es wurde zusammen mit Managern und Beratern unterschiedlicher beruflicher Spezialisierungen erarbeitet, welche aktuellen Herausforderungen unterstützt werden können. Ein oft genannter Punkt war, dass der Informations- bzw. Erfahrungsaustausch untereinander selten möglich ist. Diese Herausforderungen im Austausch gab es sowohl unter den Beratern (horizontale Kommunikation) als auch entlang der Hierarchie im Unternehmen (vertikale Kommunikation).

Es wurde eine Situation identifiziert, in der Austausch nur mit den Bürokollegen oder mit dem persönlichen Netzwerk stattfindet (vgl. für diesen Abschnitt Brown u. a. 2015, 94–108). Nicht jedes Team führt regelmäßige interne Meetings durch, um sich zu Problemfällen auszutauschen oder um gegenseitig voneinander zu lernen. Ein jährliches Event mit der kompletten Belegschaft, welches von dieser oft zum Netzwerken und Lernen genutzt wurde, wurde eingestellt. Die Berater geben zu bedenken, dass die Situationen in anderen Orten oft unterschiedlich ist (aus Gründen der Demografie, Geografie, Industrie, etc.), allerdings sollen die Probleme zwischen Büros in Städten



gleicher Größe oft sehr vergleichbar sein. Es ist nur ein Tag pro Woche für interne Meetings reserviert. Der übrige Arbeitsalltag in der Arbeitsagentur beinhaltet viele kurze Termine mit unterschiedlichen Klienten.

Neben dem Austausch gab es andere Aspekte, die Druck auf die Berater ausgeübt haben, sich mit der eigenen beruflichen Identität auseinanderzusetzen: Unter den Klienten gab es mehr Langzeitarbeitslose, die eine andere Art der Beratung benötigen, neue Berufe und höhere Erwartungen seitens der Klienten einen Job zu finden.

Ein allgemeines Problem in öffentlichen Verwaltungen ist der sogenannte *Lack of Exit*, welcher beschreibt, dass für bestimmte Dienstleistungen, wie bspw. der Empfang von Sozialhilfe, die öffentliche Verwaltung die einzige Institution ist, an die sich ein Klient wenden kann (vgl. für diesen Abschnitt Lindgren und Jansson 2013). Privatwirtschaftliche Unternehmen können sich auf der anderen Seite aussuchen, mit wem sie eine Geschäftsbeziehung eingehen wollen. Dieser Zwang sich um jeden Klienten angemessen kümmern zu müssen und Verantwortung im Sinne der Haftung zu übernehmen, verursacht Stress für die Mitarbeiter. Dies kann ein weiteres Thema für den gemeinsamen Austausch in einer Community of Practice sein.

Auf Basis dieser Befunde entschied sich das Projekt EmployID zur Implementierung einer Community of Practice Plattform, um den Mitarbeitern zu ermöglichen, selbstständig eine Community für den gegenseitigen Austausch zu schaffen. Darüber können Berater und Manager untereinander über aktuelle Herausforderungen und Ideen diskutieren. Dieser Community-Ansatz bietet einerseits den Raum für kollaborative Reflexion (siehe Abschnitt 2.1) und andererseits den Raum für das aktive Auseinandersetzen mit dem Wandel der eigenen beruflichen Identität. Aus diesem Grund ist der Kontext der Studie eine Community of Practice Plattform, die kollaborative Reflexion innerhalb einer Arbeitsagentur ermöglichen soll.

### **4.3 Methode**

In diesem Abschnitt wird der Aufbau der Studien sowie das Vorgehen für die Analyse beschrieben. Dieser Grundaufbau war in beiden Anwendungsfällen identisch.

#### **4.3.1 Studiendesign**

Zunächst wurde die Studie in einem *within-subjects*-Design mit zwei Gruppen angelegt, bei der die erste Gruppe zunächst über zwei Monate Prompts bekommt und anschließend

keine Prompts mehr bekommt (bei der anderen Gruppe vice versa) (vgl. für diesen Abschnitt Blunk und Prilla 2017b).

Die anfängliche Gruppe an Beratern, die sich auf der Community of Practice Plattform registriert hat, war kleiner als erwartet. Zusätzlich haben sich über persönliche Gespräche unter den Mitarbeitern täglich neue Personen registriert. Dadurch ergab sich eine Situation, in welcher neu registrierte Benutzer eine andere Umgebung vorgefunden haben, da sich der Umfang der Community und des Inhalts täglich ändert. Dementsprechend war die Situation für neuregistrierte Nutzer nicht vergleichbar mit der Situation für die Personen aus der anfänglichen Startgruppe. Dies ist problematisch, da so die Situation zwischen den Nutzern in den unterschiedlichen Gruppen nicht vergleichbar ist. Eine größere Anzahl von registrierten Nutzern ermöglicht mehr Aktivität in der Community (bspw. mehr Gruppen und mehr Beiträge). Somit finden neue Nutzer, die sich Monate nach Start der Community of Practice Plattform registrieren, in Bezug auf die vorhandenen Inhalte, eine stark veränderte Situation vor. Aus diesem Grund werden die Datensätze in einem Vergleich zwischen der Zeit, in den ein Berater Prompts bekommen hat und der Zeit, in der ein Berater keine Prompts bekommen hat, gegenübergestellt. Bei diesem Vergleich ist zu beachten, dass durch die lange Laufzeit der Studien eine längere Zeit mit Prompts mit einer Zeit von zwei Monaten ohne Prompts verglichen wird. Mit Ausnahme der Nutzer, die sich erst gegen Ende der Studie registriert haben, ist dieser Unterschied für die meisten Nutzer vorhanden.

Da die Literatur- und die Projektsprache überwiegend Englisch ist, wurden die Prompts initial in der englischen Sprache entworfen. Da die Mehrheit der Berater in den Arbeitsagenturen sich mit der englischen Sprache im Berufsalltag nicht wohlfühlt, wurde die Entscheidung getroffen, die Plattform und die Inhalte in die jeweilige Landessprache zu übersetzen. Zur Sicherstellung der korrekten Übersetzung der Prompts wurden die übersetzten Prompts getestet. Dazu hat eine Gruppe von drei Mitarbeitern in der Arbeitsagentur jeweils einen kleinen Text bekommen und sollte auf die Prompts eine Antwort verfassen bzw. Rückmeldung geben, wenn der Prompt unverständlich ist. Der jeweilige Übersetzer hat dann geprüft, ob die Rückmeldungen zu dem Prompt passen. In allen Fällen konnte die initiale Übersetzung übernommen werden.

#### **4.3.2 Verwendete Prompts**

Das in Abschnitt 3.5 beschriebene Softwaresystem wird genutzt, um den Benutzern die Prompts anzuzeigen. Für die Studie wurden die in Abschnitt 3.2 beschriebenen Prompts ausgewählt. Das Ziel war es, die wichtigen Aspekte von Reflexion (siehe Abschnitte 2.4

und 2.5) wie bspw. das Einbringen eigener Erfahrungen und das Vorschlagen von Lösungsvorschlägen in der Studie zu testen. Eine Liste der in der Studie genutzten Prompts befindet sich in Tabelle 4-1. Die Tabelle zeigt zu jedem Prompt-Text auch, für welche Rolle dieser sichtbar war (siehe Abschnitt 3.5.3.1.2) und wo der Prompt angezeigt werden konnte (siehe 3.5.3.1.3).

**Tabelle 4-1: Übersicht der verwendeten Prompts, wo (innerhalb einer Diskussion, neue Diskussion) diese sichtbar waren, und welche Rolle (Autor, Leser, Antwortender) im Forum den Prompt empfangen konnte und welches Ziel die Prompts aus dem Konzept abdecken (Abschnitt 3.2)**

<b>Prompt Text</b>	<b>Rolle</b>	<b>Ort</b>	<b>Ziel</b>
Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?	/	Neue Diskussion	Z1
What was unexpected for you in your daily work in the last 2 weeks? Why?	/	Neue Diskussion	Z1
Try to describe your point of view using a metaphor.	/	Neue Diskussion	Z4
What do you think about what has happened? Do you already know how you want to approach this differently in future?	/	Neue Diskussion	Z4
Try to make clear what you want to know from your colleagues	/	Neue Diskussion	Z2
What did you already try to approach the situation?	/	Neue Diskussion	Z5
What would you personally suggest from your experience? Why?	Leser, Antwortender	Innerhalb einer Diskussion	Z8
Do you agree / disagree with any of the existing comments? What would you suggest from your experience? Why?	Autor, Leser, Antwortender	Innerhalb einer Diskussion	Z6
What have you learned from this discussion so far?	Autor, Leser, Antwortender	Innerhalb einer Diskussion	Z9, Z10
Do you already have an idea how you want to tackle similar situations in future? If not, how about phrasing a new question?	Autor	Innerhalb einer Diskussion	Z11
What specific steps do you want to take to approach similar situations in future?	Autor	Innerhalb einer Diskussion	Z12
Have you already experiences after making changes in your approach? If yes, what are those experiences?	Autor	Innerhalb einer Diskussion	Z13
What do you think about how I dealt with this situation?	Vom Autor an Leser / Antwortenden	Nutzer-generierte Prompts	Z7

Prompt Text	Rolle	Ort	Ziel
What would you personally from your experience recommend as a solution to the case and why?	Vom Autor an Leser / Antwortenden	Nutzer-generierte Prompts	Z8
Have you ever had a similar experience? What happened?	Vom Autor an Leser / Antwortenden	Nutzer-generierte Prompts	Z6
What are your recent work related issues? Write down your experiences or try to help your colleagues solving their problems	Back-up (Im Falle technischer Probleme im Prompting-System, wurden diese zwei Prompts angezeigt)		
Have you ever encountered a similar situation? What happened? How did you react?			

Um eine Ermüdung des Nutzers hinsichtlich bestimmter Prompts zu vermeiden (Nückles, Hübner und Renkl 2006; van Oldenbeek u. a. 2019), zeigt das System für ein bestimmtes Ziel bzw. für die aktuell identifizierte Situation variabel unterschiedliche Prompts an. Die Grundlage dafür war der erkannte Kontext des Nutzers (die Rolle des Nutzers auf der Plattform oder innerhalb einer Diskussion). Dazu passend wurde eines der zu unterstützenden Ziele ausgewählt. Da manche Ziele mehrere Prompts haben, wurde in den Fällen ein zufälliger Prompt ausgewählt und angezeigt. Das Ziel war eine Balance zwischen Abwechslung der Prompts und ausreichender Datenmenge zu finden. Es sollten einerseits ausreichend unterschiedliche Prompts dargestellt werden, um Ermüdungserscheinung zu vermeiden, und andererseits sollten ausreichend Daten pro Prompt gesammelt werden, um diese evaluieren zu können. Für Letzteres sollte die Menge der Prompts nicht zu groß gestaltet werden, damit eine Analyse einzelner Prompts möglich ist. Wenn die Menge der Prompts sehr groß ist, werden dementsprechend mehr Beiträge benötigt, die entstanden sind, während ein Prompt angezeigt wurde.

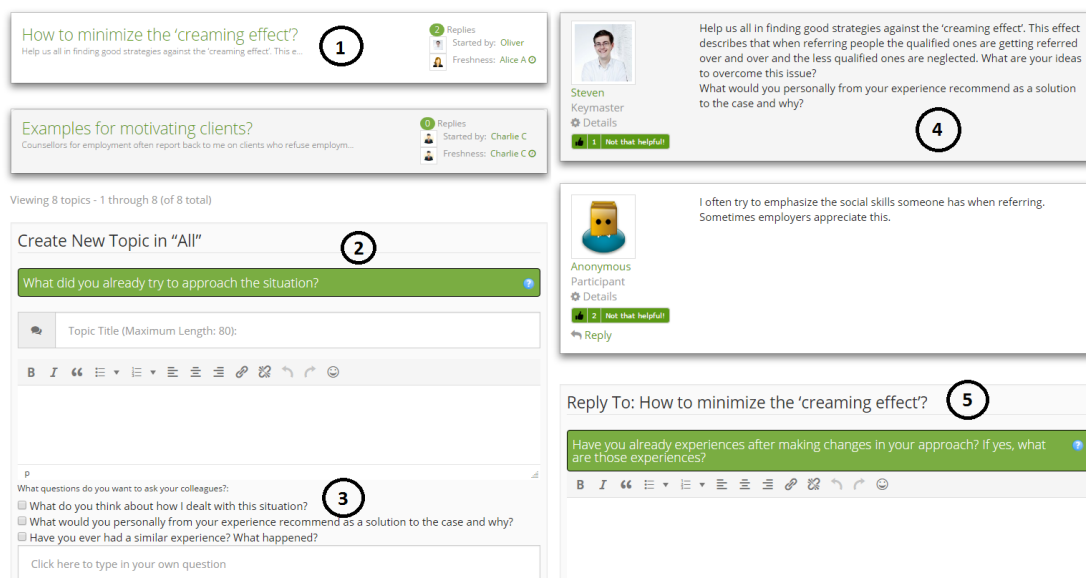


Abbildung 4-2: Prompts im Bereich der Foren (grün hinterlegt)

Die Prompts für den Autor für die Ziele Z11, Z12 und Z13 wurden erst gezeigt, sobald das Thema zwei oder mehr Antworten hatte. Ausgehend von der Literatur hilft der Erfahrungsaustausch bei der Reflexion (siehe Abschnitt 2.5). Dementsprechend soll der Autor erst nach Erfahrungsaustausch einen Plan für zukünftige Handlungen entwickeln.

Die Abbildung 4-2 zeigt die Prompts im Bereich der Foren sowohl bei der Erstellung neuer Themen (1-3) als auch bei der Beantwortung existierender Themen (4-5). Unterhalb der Liste existierender Themen (1) wird ein Prompt (2) angezeigt, der bei der Erstellung eines neuen Themas helfen soll (3). Wenn jemand eine Antwort auf ein Thema geben möchte, findet er unterhalb der Antworten (4) einen Prompt (5), der bei der Erstellung einer Antwort helfen soll.

In Abbildung 4-3 sind Prompts auf der Startseite (unten im Bild, gelb hinterlegt) dargestellt. Hier sollen die Prompts auf der Plattform dazu anregen, neue Themen zu erstellen und in diesen direkt zu reflektieren.

**EmployID – Demo of the Social Learning Platform**

Home News Groups Members Help 2 Language:

## Welcome to the ZRSZ community

Here you can get an overview what happened in the community since your last visit

### Recent News

- Kick-Off Event for the Community Platform
- New development in Ljubljana: Specializing on certain industries
- Case Study: Local Employment Partnerships in UK
- Welcome to the learning platform

### Topic of the Month

**How to minimize the 'creaming effect'?**  
Help us all in finding good strategies against the 'creaming effect'. This effect describes that when referring people the qualified ones are getting referred over and over and the less qualified ...

**Latest reply (by Alice A)**  
I often try to include one or more persons who are on paper less qualified but who were quite motivated in the appointments here at ZRSZ. This can account for a lot and I even got Feedback from ...

[Join the discussion](#)

### Recent Topics in the Forums

- My super secret topic
- How to minimize the 'creaming effect'?
- Topic of the Month Archive
- Employers making requests for candidates not eligible: What to do?
- Examples for motivating clients?

### Most useful posts in the last 30 days

- Not willing to take job - motivation?**  
There was a job vacancy at company X. If offered it to a candidat... 2 Votes
- Reply to: not-willing-to-take-job-motivation**  
The same happened to me recently. I offered a vacancy for assista... 2 Votes
- Reply to: how-to-minimize-the-creaming-effect**  
I often try to emphasize the social skills someone has when refer... 2 Votes
- Reply to: not-willing-to-take-job-motivation**  
This also happened many times to me and my colleagues here in Z. ... 1 Votes
- How to minimize the 'creaming effect'?**  
Help us all in finding good strategies against the 'creaming ef... 1 Votes

### Tags

#foreigners #motivation

### Your thought of the day

**What are your recent work related issues? Write down your experiences or try to help your colleagues solving their problems**

Post your thoughts in the:  
→ [General forum](#)

[Edit](#)

**Oliver**  
[Log Out](#)

Search Forums

Search for:  [Search](#)

Currently Online

Abbildung 4-3: Prompts auf der Startseite der Plattform (unten, gelb hinterlegt)

### 4.3.3 Vorgehen für die Analyse

Die Studie wird quantitativ analysiert, um zu überprüfen, ob die Prompts zu einer Änderung in der Reflexion geführt haben. Falls es einen Unterschied zwischen den beiden Gruppen gibt, kann angenommen werden, dass die Prompts einen Effekt hatten.

Zur genaueren Analyse soll ermittelt werden, ob ein Nutzer überhaupt einen Prompt gesehen hat. Wenn jeder Seitenaufruf so gewertet wird, als hätte der Nutzer den Prompt gesehen, dann schließt dies auch Fälle ein, in denen der Nutzer nur schnell in ein Thema reingeschaut hat und wieder zurück in die Übersicht der Themen navigiert ist, ohne dass der Prompt im Sichtfeld war. Diese Fälle verzerren die Auswertung und sollen ausgeschlossen werden. Für die Ermittlung, ob ein Prompt als gelesen gewertet werden kann, wird die folgende Heuristik gewählt (vgl. für diesen Abschnitt Blunk und Prilla 2017b): Nicht jeder angezeigte Prompt wird vom Benutzer tatsächlich gelesen. Nicht jeder Prompt, der auf der Webseite dargestellt ist, war im Sichtfeld des Nutzers. Es gibt die Möglichkeit, dass der Prompt auf der Webseite sehr weit unten dargestellt wird und der Nutzer scrollen muss, um diesen zu sehen. Bei der Analyse der Prompts werden daher nur Prompts berücksichtigt, die im Sichtfeld des Nutzers waren. Nutzer könnten rasch drüber hinweg scrollen, aber dieser Fall wurde ausgeschlossen, da der untere Abschnitt des Forums in jedem Thema gleich aussieht und der Nutzer weiter unten nichts Neues findet. Dementsprechend ist die Annahme, dass der Nutzer dieses Verhalten nur wenige Male zeigen wird und daher die Datenqualität nicht wesentlich darunter leidet. Es wäre möglich, den Nutzer bei jedem Beitrag zu fragen, ob der Prompt gesehen wurde, aber dies führt vermutlich eher zu einer persönlichen Ablehnung des Systems und ist somit der Studie nicht zuträglich. Dementsprechend ist die Annahme, dass der Nutzer den Prompt gelesen hat, sobald dieser im Sichtfeld des Nutzers war.

Zur Überprüfung, ob Beiträge Reflexion oder Elemente von Reflexion enthalten, ist eine systematische Inhaltsanalyse notwendig (siehe Abschnitt 3.4). Da der Autor der vorliegenden Arbeit die jeweiligen Landessprachen nicht spricht, haben jeweils zwei Studenten, welche die Sprache des Inhaltes des Prompts als Muttersprache sprechen, die Inhaltsanalyse durchgeführt. Dieses Vorgehen wurde so gewählt, damit keine Bedeutung im Übersetzungsprozess verloren geht. Alle Studenten hatten Vorerfahrung mit textueller Inhaltsanalyse, was ein Kriterium für die Auswahl war.

Das Schema für die Inhaltsanalyse sowie das allgemeine Vorgehen wurde in Abschnitt 3.4 beschrieben.

Die Studenten wurden von zwei der Autoren, des Kodierschemas trainiert (der Autor dieser Arbeit hat ebenfalls kodiert). Dazu haben die Studenten Testdatensätze bekommen,

die aus einem vorherigen Projekt stammen und dort bereits kodiert wurden. Die Studenten haben eine unkodierte Version der Daten bekommen und konnten so die Anwendung des Kodierschemas üben.

Für die folgenden Analysen wurden die drei neuen Codes zur Analyse von Prompts (siehe Abschnitt 3.4.2) zu einer Variable zusammengefasst, die im Folgenden *Prompt Code* genannt wird. Je nach Passung der Antwort zum Prompt wurde in der Variable die entsprechende Zahl (1, 2 oder 3) der obigen Codes vergeben. Wurde der Text ohne Prompt verfasst, wurde für den jeweiligen Beitrag in der Variable eine „0“ eingetragen. Auf diese Weise lässt sich in der Variable eine Ordinalskala über die Ziffern von eins bis drei abbilden, die eine steigende Passung des Textes zum Prompt repräsentiert. Die kodierten Textstellen hingegen sind als boolesche Variablen hinterlegt (Nominalskala) und beinhalten entweder eine „1“ (der Code ist in einer Textstelle enthalten) oder eben eine „0“, falls dies nicht der Fall ist.

#### **4.4 Anwendungsfall: Slowenien**

Im Rahmen des Forschungsprojekts EmployID war die Arbeitsagentur in Slowenien der Partner, mit dem die Community of Practice Plattform zusammen entwickelt wurde. In diesem Abschnitt wird eine kurze Übersicht zum Anwendungspartner, der geplanten Community gegeben. Hauptteil dieses Abschnitts ist die Auswertung der Studie.

##### **4.4.1 Kontext des Anwendungsfalles in Slowenien**

Bei der Arbeitsagentur in Slowenien wurde eine Erhebung durchgeführt, um die aktuelle Situation zu analysieren (vgl. im Folgenden Brown u. a. 2015, 17–21, 25–29). Basierend auf den Ergebnissen wurde die Community of Practice Plattform zur Unterstützung der vertikalen und horizontalen Kommunikation eingeführt.

Die Community of Practice Plattform basiert auf dem entwickelten Softwaresystem (siehe Abschnitt 3.5). Die Plattform wurde partizipativ mit einer kleinen Gruppe von Nutzern entwickelt, die frühzeitigen Zugriff auf eine Entwicklungsversion der Plattform bekommen haben.

Die Tabelle 4-2 zeigt eine Übersicht des zeitlichen Ablaufs der Studie mit Beginn des initialen Workshops, in dem die Plattform offiziell gestartet wurde. Es ist zu beachten, dass ursprünglich zwei Gruppen in einem within-subjects Design geplant waren, was aufgrund der niedrigen Zahl der anfänglichen Nutzer, der sich konstant verändernden Umgebung, der kontinuierlich steigenden Nutzerzahl, und der längeren Laufzeit anders

ausgewertet wurde. Bis zum Ende des Datenerfassungszeitraums wurden neu-registrierte Nutzer in die Studie aufgenommen.

**Tabelle 4-2: Ablauf der Studie in Slowenien**

<b>Datum</b>	<b>Event</b>
04.02.2016	Start der Community of Practice in einem Workshop mit 18 Personen.
04.04.2016	Nutzer im within-subjects Design wechseln die Kondition (von Prompt zu Nicht-Prompt und von Nicht-Prompt zu Prompt).
06.06.2016	Für die Nutzer, die seit dem initialen Workshop registriert sind, endet die ursprüngliche Laufzeit der Studie.
13.06.2016	Ab diesem Zeitpunkt sehen alle Nutzer, deren Registrierungszeitpunkt älter als die zwei mal zwei Monate für den Zeitraum der Studie ist, immer Prompts. Nutzer, die vor diesem Zeitpunkt länger als der Zeitraum der Studie registriert waren, haben keine Prompts gesehen.
07.07.2016	Interviews in Slowenien mit 17 Nutzern. Zu diesem Zeitpunkt waren bereits mehr als 100 Nutzer registriert. Die Teilnahme an den Interviews war freiwillig.
28.02.2018	Letzter Tag, an dem Nutzungsdaten gesammelt wurden, bevor mit der Analyse für den Projektabschluss begonnen wurde.

Eine Auflistung, welche Faktoren die Entwicklung der Community begünstigt oder behindert haben, findet sich in (Blunk und Prilla 2017a). Eine Darstellung, wie sich die Community hinsichtlich der Zahl der Mitglieder und der Themengruppen entwickelt hat, findet sich in (EmployID 2018, 72–81).

Für das allgemeine Vorgehen in der Analyse siehe Abschnitt 4.3.3. Im Folgenden werden nur spezifische Analyseschritte für diesen Anwendungsfall samt den Ergebnissen beschrieben.

Allgemein muss für diesen Anwendungsfall hervorgehoben werden, dass für die vorliegende Dissertation die deskriptive bzw. quantitative Analyse basierend auf der Produktivdatenbank durchgeführt wurde. Für die Inhaltsanalyse (qualitative Analyse) wurde in 02/2018 der damalige Stand an Beiträgen ausgewählt und inhaltlich kodiert. Da die Nutzer und Moderatoren seitdem mehrere Beiträge gelöscht haben, fand die Inhaltsanalyse auf einem leicht unterschiedlichen Datensatz statt. Ein Problem entsteht, sobald ein Nutzer seinen eigenen Beitrag entfernen möchte, da in diesem Fall auch die dazugehörigen Antworten ebenfalls gelöscht werden. Nutzer waren ab einem Zeitpunkt von wenigen Stunden nach dem ursprünglichen Schreiben nicht mehr in der Lage Beiträge zu bearbeiten bzw. zu löschen. Moderatoren konnte in dem Fall aber helfen. Trotz Hinweise seitens der Forscher Beiträge zu Forschungszwecken nicht vollständig zu löschen, sondern nur als gelöscht zu markieren, damit sie für die Nutzer nicht mehr auftauchen, wurden Beiträge dennoch vollständig gelöscht. Die Moderatoren hatten diese



Berechtigungen, da eine der Zielsetzungen im Projekt war, dass die Moderatoren langfristig (über das Projektende hinaus) die Plattform allein betreuen sollen, wofür entsprechende Berechtigungen notwendig sind. Löschpflichten in Datenschutzverordnungen sind ein weiterer Grund, warum die Moderatoren die Möglichkeit benötigen, gezielt Beiträge löschen zu können.

#### 4.4.2 Deskriptive Analyse des Anwendungsfalles in Slowenien

Die Community hat im Verlauf des Datenerhebungszeitraums einen nennenswerten Anteil der Berater der Arbeitsagentur in Slowenien (ca. 900 Mitarbeiter) angezogen (siehe Tabelle 4-3). Es ist nur ein Anteil von ca. zehn Prozent der Nutzer durch das Verfassen von Beiträgen (Themen oder Antworten) aktiv geworden. Der Wert ist deutlich kleiner als die in der Literatur genannten 25-35 % (Wenger, McDermott und Snyder 2002, 56). Ein Anteil von ca. 13 % der Nutzer ist als inaktiv zu klassifizieren, was bedeutet, dass diese Personen weder Beiträge geschrieben noch Beiträge im Forum gelesen haben. Eine oberflächliche Analyse der Log-Dateien ergab, dass sich diese Nutzer die Startseite, die Übersicht der Gruppen und hauptsächlich den persönlichen Profilbereich angeschaut haben. Dies kann darauf hindeuten, dass aus dem Design der Plattform nicht eindeutig hervorgeht, wo sich der Diskussionsbereich befindet. In der Community waren insgesamt drei Moderatoren tätig, von denen zwei seit Beginn der Community tätig waren.

**Tabelle 4-3: Aktivität der Benutzer und Moderatoren in der Community  
in Slowenien zwischen 02.2016 und 02.2018 ohne Gruppen für E-Coaching oder DrogArt**

<b>Aktivität</b>	<b>Nutzer</b>	<b>Moderatoren</b>
Anzahl registrierte Personen	466	3
Anzahl aktive Personen (lesen und schreiben)	43	/
Anzahl passive Personen (nur lesen)	361	/
Anzahl inaktive Personen (nur registriert)	62	/
Anzahl Lese-Events	6006	1720
Anzahl geschriebene Themen	40	58
Anzahl geschriebene Antworten	138	40

Die Daten in Tabelle 4-3 umfassen nicht alle Werte: Es wurden zwei Nutzer ausgeschlossen die mit 24 (anonymisierte Nutzer-ID u/4996) und 21 (anonymisierte Nutzer-ID u/2798) Beiträgen deutlich mehr Beiträge als die übrigen Nutzer hatten. Durch diese sehr hohe Anzahl an Beiträgen lässt sich in der Analyse nicht zweifelsfrei ermitteln, ob Unterschiede zwischen den Gruppen an den Prompts liegen oder durch andere Faktoren hervorgerufen werden. Die Personen, die Beiträge geschrieben haben, haben durchschnittlich 4,23 Beiträge geschrieben (SD=4,94; Mdn.=2,5). Diese Werte zeigen, dass der Großteil der Nutzer wenig Beiträge geschrieben und hat und wenige Nutzer sehr

viele Beiträge geschrieben haben, was für Communities of Practice normal ist (siehe Abschnitt 2.1).

Bei allen folgenden Analysen wurde die Aktivität der Nutzer und Moderatoren in zwei Gruppen ausgeschlossen: Die erste Gruppe („*DrogArt*“) beinhaltete ein spezielles Event, bei dem ein externer Berater hinzugezogen wurde, der Fragen zu Fällen beantwortet hat, in denen die Arbeitssuchenden bspw. Drogenprobleme haben. Dieses Event wurde sehr gut besucht, hat aber die Werte in der Studie verzerrt. Die zweite Gruppe umfasst den Diskussionsbereich für einen online durchgeführten Peer-Coaching-Kurs, in dem die Nutzer als Hausaufgabe Diskussionsbeiträge erstellen sollten. Dieser Kurs erfreut sich sehr großer Beliebtheit und lief in einem Zeitraum von ca. sechs Wochen. Insbesondere durch die Hausaufgabe wurden die Daten in der Studie ebenfalls stark verzerrt. Aus diesen Gründen wurde die Aktivität in den beiden genannten Gruppen ausgeschlossen.

#### **4.4.3 Quantitative Analyse des Anwendungsfalles in Slowenien**

In Tabelle 4-4 werden die ursprünglichen Studienzeiträume für die Nutzer und Moderatoren gezeigt. Die Moderatoren in diesem Anwendungsfall waren die Personen, die seitens der Arbeitsagentur im Forschungsprojekt mitgearbeitet haben. Innerhalb der Organisation waren diese Personen für Projektarbeit zuständig und sie waren nicht als Berater tätig. Mit den ursprünglichen Studienzeiträumen ist gemeint, dass alle Aktivitäten, die in den ersten vier Monaten ab Registrierung bei einem Nutzer angefallen sind (jeweils zwei Monaten mit und zwei Monaten ohne Prompts), abgebildet werden. Zunächst ist der Tabelle 4-4 zu entnehmen, dass die Lese-Aktivitäten der Nutzer in den beiden Zeiträumen ähnlich sind. Bei den Schreib-Aktivitäten zeigt sich, dass in der Prompt-Kondition mehr Beiträge aber knapp weniger Antworten geschrieben wurden. Wenn man eine Quote aus der Summe der jeweiligen Lese- und Schreib-Aktivitäten errechnet, dann zeigt sich, wie viele Beiträge es pro Lesezugriff auf die Liste der Themen oder auf die Antworten innerhalb eines Themas (siehe Abbildung 4-2 (1) und (4)) gab. Bei Nutzern scheint die Anwesenheit von Prompts mehr neue Themen pro Lesezugriff zu implizieren. Bei der Quote für Antworten sinkt scheinbar die Anzahl neuer Antworten pro Lesezugriff, wenn Prompts vorhanden sind.

Aufseiten der Moderatoren ist die Anzahl der Lesezugriffe in beiden Perioden ausgeglichen. Bei den Beiträgen sind die Zahlen in den beiden Konditionen stark unterschiedlich und implizieren, dass die Moderatoren bei der Anwesenheit von Prompts deutlich weniger Beiträge pro Lesezugriff schreiben.

**Tabelle 4-4: Ursprünglicher Zeitraum, ohne die ausgeschlossenen Nutzer (anonymisierte Nutzer-IDs u/4996 und u/2798), ohne die Aktivitäten in den Gruppen für Peer-Coaching-Kurs und für schwervermittelbare Arbeitssuchende (DrogArt)**

	Nutzer		Moderatoren	
	Keine Prompts	Prompts	Keine Prompts	Prompts
Alle Lese-Events	2649	2574	865	703
Lese-Events Forum	1194	1036	489	392
Lese-Events Beiträge	1455	1538	376	311
Erstellte Themen	8	13	23	11
Erstellte Antworten	34	31	16	1
Lese-Schreib-Quote Forum	0,67 %	1,25 %	4,70 %	2,81 %
Lese-Schreib-Quote Beitrag	2,34 %	2,02 %	4,26 %	0,32 %

Wenn man den kompletten Zeitraum von zwei Jahren betrachtet (02/2016 bis 02/2018), dann ist die Phase, in der die Nutzer Prompts gesehen haben, für den Großteil der Nutzer wesentlich länger. Aus Tabelle 4-5 ist ersichtlich, dass in diesem Zeitraum wesentlich mehr Aktivität lesender und schreibender Art erfolgt ist. Die Verteilung der Quoten zu dem ursprünglichen Studienzeitraum ist weiterhin gleich und impliziert, dass die Prompts zu mehr Beiträgen pro Lesezugriff zu führen scheinen und die Prompts die Rate von Antworten zu behindern scheinen. Auf Seite der Moderatoren ist das Verhältnis ebenfalls gleichgeblieben. Der einzige Unterschied zu Tabelle 4-4 ist, dass die Moderatoren in dem verlängerten Zeitraum überproportional viele Antworten verfasst haben. Dennoch ist die Quote für Moderatoren gleichgeblieben und impliziert, dass bei Moderatoren die Anwesenheit von Prompts das Erstellen von neuen Themen und Antworten eher verringert als begünstigt. Die Daten, welche Prompts bei den jeweiligen Beiträgen sichtbar waren, wurden aufgrund eines technischen Problems nicht alle erfasst und bleiben daher von den weiteren Analysen ausgeschlossen. Dementsprechend findet die weitere Analyse nur auf den Daten des ursprünglichen Experimentzeitraums statt.

**Tabelle 4-5: Kompletter Zeitraum, ohne die ausgeschlossenen Nutzer (anonymisierte Nutzer-IDs u/4996 und u/2798), ohne Peer-Coaching-Kurs, ohne DrogArt**

	Nutzer		Moderatoren	
	Keine Prompts	Prompts	Keine Prompts	Prompts
Alle Lese-Events	2790	6560	901	2348
Lese-Events Forum	1241	2801	509	1434
Lese-Events Beiträge	1549	3759	392	914
Erstellte Themen	9	27	23	38
Erstellte Antworten	36	58	19	21
Lese-Schreib-Quote Forum	0,73 %	0,96 %	4,52 %	2,65 %
Lese-Schreib-Quote Beitrag	2,32 %	1,54 %	4,85 %	2,30 %

Für beide Tabellen, Tabelle 4-4 und Tabelle 4-5, ist festzuhalten, dass die Unterschiede zwischen der Nicht-Prompt-Kondition und der Prompt-Kondition minimal sind und teilweise nur 0,25 % umfassen. Daher dürfen die Ergebnisse an dieser Stelle nicht überinterpretiert werden.

In Tabelle 4-6 wird eine Übersicht gezeigt, wie oft, welcher Prompt vom Server abgerufen wurde, wie häufig dieser Prompt dann anschließend im Sichtfeld des Nutzers war, und wie oft daraus ein neuer Beitrag entstanden ist. In der Prompt-Log-Tabelle sind Daten gelöschter Beiträge aus dem ursprünglichen Zeitraum aufgeführt. Da es in dieser Tabelle um das Verhalten geht, wie oft ein Prompt gesehen wurde und zu einem Beitrag führte, wurden diese Logdaten beibehalten.

Im Bereich der Prompts, die einen neuen Beitrag begünstigen bzw. zur Reflexion beeinflussen sollen, lässt sich feststellen, dass zwei der Prompts sehr häufig (ca. 175x) gesehen wurden, aber nie zu einem Beitrag geführt haben („*Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?*“ und „*What was unexpected for you in your daily work in the last 2 weeks? Why?*“; Z1). Diese beiden Prompts scheinen bei Nutzern nicht auf positive Resonanz zu stoßen und können daher in künftigen Studien ausgeschlossen werden. Eine Vermutung ist, dass zwar Reflexion durch Unerwartetes ausgelöst wird (siehe Abschnitt 2.4), dies aber kein Ansatz für eine Frage bzw. ein Prompt ist. Eine andere Vermutung ist, dass Nutzer die Plattform zu dem Zeitpunkt nicht in der Form nutzen, als dass sie für einen Prompt dieser Art offen sind. Die anderen Prompts haben jeweils zu einer kleinen Anzahl an Beiträgen geführt, aber keiner dieser Prompts sticht überdurchschnittlich hervor. Daher können die übrigen Prompts in dieser Kategorie für künftige Studien beibehalten werden.

Bei den Prompts, die einen Beitrag innerhalb einer Diskussion begünstigen bzw. beeinflussen sollen, gibt es im Wesentlichen zwei Cluster. Die Prompts für die Ziele Z6, Z8, Z9 bzw. Z10 wurden häufig gesehen und traten in zwei Prozent der Fälle zusammen mit Beiträgen auf. Diese Gruppe von Prompts kann für weitere Studien beibehalten werden. Die anderen drei Prompts sind alle an den Autor eines Beitrags adressiert und versuchen diesen dazu zu bewegen konkrete Schritte für die Zukunft zu planen („*Do you already have an idea how you want to tackle similar situations in future? If not, how about phrasing a new question?*“ (Z11) und „*What specific steps do you want to take to approach similar situations in future?*“ (Z12)) oder einen neuen Reflexionszyklus anzustoßen („*Have you already experiences after making changes in your approach? If yes, what are those experiences?*“; Z13). Diese Gruppe von Prompts sollte in weiteren Studien beibehalten werden, obwohl einer der Prompts nicht in Zusammenhang mit

Beiträgen aufgetaucht ist, weil die Datenmenge zu klein ist, als dass eine realistische Aussage zur Wirkung der Prompts getätigt werden kann.

Nutzer hatten die Möglichkeiten eigene Fragen in ihren Themen zu definieren (siehe Abbildung 4-2 (3)), welche die systemgenerierten Prompts überschrieben haben. Diese Prompts werden in dieser Arbeit als nutzergenerierte Prompts bezeichnet. In der Benutzeroberfläche sind drei verschiedene vorgefertigte Fragen vorhanden, von denen Nutzer keine, eine oder mehrere auswählen können, oder eine eigene selbstverfasste Frage anhängen können. Aus der Tabelle 4-6 geht hervor, dass Nutzer all diese Möglichkeiten genutzt haben. Ein Nutzer hat vorgefertigte Fragen mit einer eigenen Frage kombiniert, wobei in dem Fall keine Antworten entstanden sind. Hier ist ein größerer Datenbestand notwendig, um diese Form von Prompts besser einschätzen zu können. Einzig der dritte vordefinierte Prompt wurde ca. 90x gesehen und trat in drei Prozent der Fälle mit einem Beitrag auf, was über den Werten für andere Antworten liegt. Die Kombination aller drei Fragen schien ebenfalls häufig mit Antworten aufzutreten.

Der letzte Block in Tabelle 4-6 zeigt die rein nutzergenerierten Fragen, die nicht in Verbindung mit vordefinierten Fragen gestellt wurden. Hier trat die Mehrheit der Prompts nicht in Zusammenhang mit Antworten auf. Bei den Prompts, die in Zusammenhang mit Antworten auftraten, ist die Häufigkeit, in der diese gesehen wurden, sehr niedrig. Daher lässt sich keine eindeutige Aussage treffen.

Tabelle 4-7 zeigt, wie oft ein Prompt an einem Ort (Mögliche Orte: bei Erstellung von Themen, innerhalb von Themen, auf der Startseite) in der Community of Practice zusammen mit einem Beitrag auftrat. Für neue Themen lässt sich aufseiten der Nutzer ableiten, dass in 63 % der Seitenbesuche ein Prompt im Sichtfeld des Nutzers war (während der Nutzer in der Prompt-Kondition war). Aus diesen Seitenbesuchen heraus, haben die Nutzer zu drei Prozent einen Beitrag verfasst. Moderatoren haben die Prompts für neue Themen ca. fünf Prozent häufiger gesehen, wobei diese weniger in Zusammenhang mit neuen Beiträgen aufgetaucht sind. Da die Prompts für neue Themen am Ende der Liste aller Themen sind (siehe Abbildung 4-2 (2)) kann dies bedeuten, dass die Moderatoren häufiger bis ans Ende der Liste aller Themen scrollen. Dies bedeutet aber nicht, dass sie dann häufiger einen Beitrag schreiben.

Bei den Antworten sieht dies anders aus: Dort haben Nutzer Prompts bei 73 % der Seitenbesuche gesehen und bei diesen Besuchen wurde in 2,5 % der Fälle eine Antwort verfasst. Hier liegt die Quote knapp unter der Quote von neuen Beiträgen. In Beiträgen haben Moderatoren im Vergleich zu Nutzern seltener einen Prompt gesehen und statistisch wesentlich seltener eine neue Antwort geschrieben. Die Moderatoren in diesem

Anwendungsfall haben weniger Antworten geschrieben, damit die Antwort nicht sofort von den Beratern als die einzig gültige Antwort akzeptiert wird und die Diskussion dadurch zum Erliegen kommt. Dementsprechend ist dieser Wert nachvollziehbar.

Nutzergenerierte Prompts waren fast gleich oft, wie die übrigen Prompts für Antworten, im Sichtfeld des Nutzers. Hier gab es allerdings in 7,4 % der Fälle eine Antwort. Damit ist die Quote knapp dreimal so groß, wie bei den üblichen Antworten, was impliziert, dass nutzergenerierte Prompts eher zu Antworten führen können. Die naheliegende Interpretation ist, dass die Wortwahl der Prompts für die Nutzer vertrauter ist, weil sie ein Kollege verfasst hat, und dass der Prompt in der Situation spezifischer war als die generischen Prompts für Reflexion. Ein weiterer Grund ist, dass der vom Nutzer eingegebene Text sowohl im Thema als auch im Prompt auftauchte und der Nutzer daher an mehreren Stellen den Text sehen konnte. Aus dieser Tabelle heraus wirken die nutzergenerierten Prompts sehr vielversprechend und sollten in künftigen Studien näher beleuchtet werden. Moderatoren haben keine Antworten auf nutzergenerierte Fragen verfasst. Die Statistik zeigt, dass Moderatoren die nutzergenerierten Fragen häufiger gesehen haben als die regulären Prompts. Da beide an der gleichen Stelle angezeigt wurden, gibt es hierfür keine logische Begründung.

Aus der Tabelle 4-7 lässt sich entnehmen, dass die Prompts auf der Startseite kein Anlass für Nutzer waren, neue Themen zu verfassen. Das Element auf der Startseite scheint also keine Wirkung zu haben. Die Prompts auf der Startseite waren bei ca. 55 % aller Seitenbesuche zu sehen, was ein niedrigerer Wert als für die anderen Prompts ist. Vermutlich haben die Nutzer die Startseite eher dazu genutzt, um einen Überblick über die neuen Themen und Antworten, und übrigen Informationen in der Community of Practice zu bekommen, als den Ort für Ideen in neuen Beiträge zu nutzen. Dementsprechend wird die Startseite vermutlich eher dazu genutzt in den Forenbereich zu wechseln, als von dort neue Themen zu verfassen. Es entspricht vermutlich nicht dem mentalen Modell der Nutzer auf der Startseite zu neuen Beiträgen inspiriert zu werden und von dort direkt zu der Erstellung eines Themas in den Forenbereich zu wechseln. Dieses Element sollte somit überarbeitet werden.

**Tabelle 4-6: Übersicht welcher Prompt in Slowenien wie oft gesehen wurde und wie oft ein Beitrag geschrieben wurde (sortiert nach Ort).**

<b>Ziel</b>	<b>Prompt-Text</b>	<b># Prompt dargestellt</b>	<b>% Prompts im Sichtfeld</b>	<b>#Beiträge</b>	<b>% Beiträge bei sichtbaren Prompts</b>	<b>Ort</b>
Z1	Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?	331	57,40 %	0	0 %	Neue Disk.
Z1	What was unexpected for you in your daily work in the last 2 weeks? Why?	375	56,27 %	0	0 %	Neue Disk.
Z4	Try to describe your point of view using a metaphor.	187	65,24 %	6	5 %	Neue Disk.
Z4	What do you think about what has happened? Do you already know how you want to approach this differently in future?	212	65,09 %	5	4 %	Neue Disk.
Z2	Try to make clear what you want to know from your colleagues.	392	67,35 %	6	2 %	Neue Disk.
Z5	What did you already try to approach the situation?	420	61,19 %	4	2 %	Neue Disk.
Z8	What would you personally suggest from your experience? Why?	316	67,72 %	4	2 %	Innerhalb Disk.
Z6	Do you agree / disagree with any of the existing comments? What would you suggest from your experience? Why?	328	76,52 %	4	2 %	Innerhalb Disk.
Z9, Z10	What have you learned from this discussion so far?	1014	73,67 %	18	2 %	Innerhalb Disk.
Z11	Do you already have an idea how you want to tackle similar situations in future? If not, how about phrasing a new question?	8	75,00 %	1	17 %	Innerhalb Disk.
Z12	What specific steps do you want to take to approach similar situations in future?	8	62,50 %	1	20 %	Innerhalb Disk.
Z13	Have you already experiences after making changes in your approach? If yes, what are those experiences?	9	77,78 %	0	0 %	Innerhalb Disk.

<b>Ziel</b>	<b>Prompt-Text</b>	<b># Prompt dargestellt</b>	<b>% Prompts im Sichtfeld</b>	<b>#Beiträge</b>	<b>% Beiträge bei sichtbaren Prompts</b>	<b>Ort</b>
Z7	What do you think about how I dealt with this situation?	5	100,00 %	0	0,00 %	QS-Plugin #1
Z7, Z8	What do you think about how I dealt with this situation? What would you personally from your experience recommend as a solution to the case and why?	24	95,83 %	5	21,74 %	QS-Plugin #1 + #2
Z6, Z7, Z8	What do you think about how I dealt with this situation? What would you personally from your experience recommend as a solution to the case and why? Have you ever had a similar experience? What happened?	57	71,93 %	3	7,32 %	QS-Plugin #1 + #2 + #3
/	What do you think about how I dealt with this situation? What would you personally from your experience recommend as a solution to the case and why? Have you ever had a similar experience? What happened? Should the customer in this case be obliged to register in the CI register? <sup>22</sup>	3	100,00 %	0	0,00 %	QS-Plugin #1 + #2 + #3 + Nutzer (letzter Satz)
Z6, Z7	What do you think about how I dealt with this situation? Have you ever had a similar experience? What happened?	27	85,19 %	0	0,00 %	QS-Plugin #1 + #3
Z8	What would you personally from your experience recommend as a solution to the case and why?	2	100,00 %	0	0,00 %	QS-Plugin #2
Z6, Z8	What would you personally from your experience recommend as a solution to the case and why? Have you ever had a similar experience? What happened?	54	72,22 %	1	2,56 %	QS-Plugin #2 + #3
Z6	Have you ever had a similar experience? What happened?	122	76,23 %	3	3,23 %	QS-Plugin #3

<sup>22</sup> Übersetzung aus dem Slowenischen mit Google Translate (<https://translate.google.com/>; letzter Zugriff am 12.08.2019)



<b>Ziel</b>	<b>Prompt-Text</b>	<b># Prompt dargestellt</b>	<b>% Prompts im Sichtfeld</b>	<b>#Beiträge</b>	<b>% Beiträge bei sichtbaren Prompts</b>	<b>Ort</b>
/	Does it have any thyroid problems? I invite you to exchange experiences <sup>22</sup>	19	42,11 %	1	12,50 %	Nutzer
/	Do you find this Manual interesting and useful for our work? <sup>22</sup>	2	0,00 %	0	0,00 %	Nutzer
/	Are you now approached to identify the customer's motives and adjust the advisory process, and if so, how do you do it? <sup>22</sup>	11	72,73 %	2	25,00 %	Nutzer
/	Please suggest topics that you find particularly interesting to post them as the theme of the month for further discussion. <sup>22</sup>	12	66,67 %	0	0,00 %	Nutzer
/	What are the things you have recently encountered at work? Write down your experiences or help your colleagues solve their problems. <sup>22</sup>	48	70,83 %	0	0,00 %	Nutzer
/	Have you ever been in a similar situation? What happened? How did you react? <sup>22</sup>	68	67,65 %	0	0,00 %	Nutzer
Gesamt Abrufe/Beiträge und Durchschnitt gesehen/Beitrag verfasst		4054	67,71 %	64	2,33 %	

**Tabelle 4-7: Sichtbarkeit pro Prompt-Kategorie und Ort in Slowenien**

<b>Ort</b>	<b>Nutzer</b>			<b>Moderatoren</b>		
	Anteil im Log	Anteil wie oft sichtbar	Anteil wie oft Beitrag geschrieben von sichtbar	Anteil im Log	Anteil wie oft sichtbar	Anteil wie oft Beitrag geschrieben von sichtbar
In der Liste der Themen: Prompt vom Server abgerufen	36,73 %	36,73 %		31,98 %	31,98 %	
In der Liste der Themen: Prompt im Sichtfeld	61,37 %	63,27 %		66,83 %	68,02 %	
Ein neues Thema wurde erstellt	1,90 %		3,00 %	1,19 %		1,75 %
Während des Lesens eines Beitrags: Prompt vom Server abgerufen	26,38 %	26,38 %		30,33 %	30,33 %	
Während des Lesens eines Beitrags: Prompt im Sichtfeld	71,79 %	73,62 %		69,07 %	69,67 %	
Eine neue Antwort wurde erstellt	1,83 %		2,49 %	0,60 %		0,86 %
Startseite: Prompt vom Server abgerufen	44,76 %			40,29 %		
Startseite: Prompt im Sichtfeld	55,24 %			59,71 %		
Vom Nutzer überschriebener Prompt vom Server abgerufen	25,91 %	25,91 %		21,54 %	21,54 %	
Vom Nutzer überschriebener Prompt im Sichtfeld	68,61 %	74,09 %		78,46 %	78,46 %	
Eine neue Antwort wurde erstellt	5,47 %		7,39 %	0,00 %		

#### 4.4.4 Qualitative Analyse des Anwendungsfalles in Slowenien

Nach der vorherigen quantitativen Analyse soll an dieser Stelle eine qualitative Analyseperspektive für die Studie gewählt werden. Eines der Ziele der Inhaltsanalyse war die Analyse des Reflexionsverhaltens der Mitglieder. Die Inhaltsanalyse war nicht ausschließlich auf diese Prompting-Studie bezogen. Aus diesem Grund sind die Beiträge der Gruppe mit Problemen von schwer vermittelbaren Nutzern (DrogArt-Gruppe) und die beiden zuvor nicht einberechneten Nutzer (anonymisierte Nutzer-IDs u/4996 und u/2798) eingeschlossen. Die Beiträge aus dem Peer-Coaching-Kurs sind weiterhin ausgeschlossen. Dieser Datensatz umfasste 277 Themen und Antworten, wovon 56 Beiträge mit einem Prompt assoziiert sind. Diese niedrigere Zahl an Beiträgen mit Prompts liegt u. a. daran, dass manche Nutzer / Moderatoren (eigene) Beiträge entfernt haben. Zudem sind in die Inhaltsanalyse nicht alle Themen eingeflossen, da es zu dem Zeitpunkt mehrere Themen ohne Antworten gab. Da das Ziel der Inhaltsanalyse die Analyse des Reflexionsverhaltens ist, waren reine Diskussionsanfänge in Form von Themen ohne Antworten nicht relevant.

In dieser Analyse ist der komplette Datenerfassungszeitraum abgebildet (und nicht der Zeitraum der ursprünglichen Studiendauer). Die im Datenerfassungszeitraum geschriebenen Themen und Antworten wurden inhaltlich mit dem reflexionsbezogenen Kodierschema kodiert (siehe Abschnitt 3.4). Das durchschnittliche Krippendorffs-Alpha für die Kodierung liegt bei .83 (min. .72, max. 1), was deutlich über dem von Krippendorff empfohlenen Mindestwert von  $\alpha \geq .667$  liegt und überwiegend über dem empfohlenen Wert zur Interpretation von  $\alpha \geq .800$  liegt (Krippendorff 2004).

Die Beiträge wurden mit den drei Prompt-bezogenen Codes inhaltlich kodiert (siehe Abschnitt 3.4.2). Das durchschnittliche Krippendorffs-Alpha für diese Codes liegt bei .96 (min. .91 und max. 1).

Für die Analyse wurden nur Codes berücksichtigt, die von beiden Kodierenden gleichermaßen zugewiesen wurde. Dementsprechend basiert die Analyse auf vollständiger Übereinstimmung. Deshalb ist die Zahl der analysierten Prompts, die präsent waren, als ein Text verfasst wurde, geringer als die Zahl in Tabelle 4-6.

Tabelle 4-8 zeigt, wie oft die Prompt-bezogenen Codes der Inhaltsanalyse von beiden Kodierenden zugewiesen wurden und wie es Beiträge ohne Prompts gab. Generell lässt sich an dieser Stelle feststellen, dass es nur 24 Beiträge gibt, bei denen die Kodierenden festgestellt haben, dass der Prompt die Texte teilweise oder vollständig beeinflusst haben kann. Anteilig gesehen, sind dies 46 % der Beiträge mit Prompts, die inhaltlich kodiert

wurden. Aufgrund dieser kleinen Zahlen wurden die beiden Codes, nach denen die Beiträge teilweise oder vollständig von Prompts beeinflusst sein können, zu einer Variable kombiniert, die *Einfluss* genannt wird. Dies ist eine boolesche Variable, die entweder zeigt, dass es einen Einfluss des Prompts auf den Beitrag geben kann (1) oder eben nicht (0).

**Tabelle 4-8: Häufigkeit des Einflusses eines Prompts auf ein Thema (Slowenien)**

Prompt-Einfluss auf den Beitrag	Häufigkeit	Gruppierung
Kein Prompt bei Erstellung des Beitrags involviert	225	Kein Einfluss
Code: User did NOT respond to prompt	28	
Code: User responded in PARTS to the prompt	8	Einfluss vorhanden
Code: User responded fully to the prompt	16	

Im Folgenden sind alle Korrelationen mit \* für eine signifikante Beziehung auf dem 0,05 Level und \*\* für das 0,01 Level gekennzeichnet. Eine Korrelationsanalyse dieser Einflussvariable zeigt, dass Texte, die von Prompts beeinflusst sein können, eine schwache Korrelation mit dem Code für erfahrungsbasierte Lösungen (SUG\_EXP) haben (siehe Tabelle 4-9). Andere Effekte auf die Texte scheint allgemein nicht feststellbar zu sein.

**Tabelle 4-9: Allgemeine Korrelation Einfluss und die reflexionsbezogenen Codes (n=277 Beiträge, bei denen in 24 Beiträgen ein Einfluss kodiert wurde; 66x Code SUG\_EXP kodiert)**

	SUG_EXP
<b>Einfluss</b>	
Pearson Correlation	,129*
Sig. (2-tailed)	0,032

Aufgrund der Ergebnisse aus dem vorherigen Abschnitt hinsichtlich den höheren Antwortraten bei nutzergenerierten Prompts, wurden diese ebenfalls einer Analyse unterzogen. Die Tabelle 4-10 zeigt, dass die Antworten bei nutzergenerierten Prompts ebenfalls mehr lösungsbasierte Erfahrungen (SUG\_EXP) ausweisen. Hier ist die Korrelation stärker als bei der Analyse des gesamten Datensatzes in Tabelle 4-9. Zusätzlich ist eine negative Korrelation mit dem Code für Fragen nach Interpretation (Q\_INT) vorhanden. Dies impliziert, dass nutzergenerierte Prompts die Antwortenden dazu verleiten mehr Lösungsvorschläge auf Basis eigener Erfahrungen beizusteuern, was ein wichtiger Faktor in der kollaborativen Reflexion ist (siehe Abschnitt 2.5). Andererseits werden weniger Fragen zu einer Interpretation gestellt, welche ein wichtiges Hilfsmittel sind, um Reflexion anzuregen (siehe Abschnitt 2.10) (bspw. Daudelin 1996). Eine Interpretation für die sinkende Anzahl von Fragen nach Interpretation ist, dass nutzergenerierte Prompts die Konversation effektiver in eine Richtung lenken im Vergleich zu Fällen, in denen dort kein Text vorhanden ist.

**Tabelle 4-10: Einfluss von nutzergenerierten Prompts (n=30 Beiträge, bei denen in 19 Beiträgen ein Einfluss kodiert wurde; 8x Code SUG\_EXP und 13x Code Q\_INT kodiert)**

		SUG_EXP	Q_INT
<b>Einfluss</b>	Pearson Correlation	,480**	-,526**
	Sig. (2-tailed)	0,007	0,003

Aus Tabelle 4-6 ging hervor, dass der Prompt, der danach fragt, ob der Empfänger in der Diskussion etwas gelernt hat („*What have you learned from this discussion so far?*“ Z9, Z10) oft in Zusammenhang mit Antworten auftauchte. Eine Korrelationsanalyse ergab, dass Antworten, die entstanden sind als dieser Prompt präsent war, in Zusammenhang mit Double Loop Learning (D\_LOOP) und Fragen nach weiteren Informationen auftauchen (Q\_INF). Hier gilt es zu beachten, dass in dieser Analyse (siehe Tabelle 4-11) nur sehr wenige Dokumente (n=33) enthalten sind, von denen aber laut Tabelle 4-6 17 Dokumente mit dem Prompt entstanden sind. Dieses Ergebnis scheint zu implizieren, dass dieser Prompt Lernen begünstigen kann. Andernfalls wurden die entsprechenden Codes nur selten vergeben, was das Ergebnis an dieser Stelle verzerrt.

**Tabelle 4-11: Korrelation vom lernbezogenen Prompt (n=33 Beiträge, bei denen in 2 Beiträgen ein Einfluss kodiert wurde; 1x Code D\_LOOP und 2x Code Q\_INF kodiert)**

		D_LOOP	Q_INF
<b>Einfluss</b>	Pearson Correlation	,696**	,468**
	Sig. (2-tailed)	0,000	0,006

#### 4.4.5 Analyse der Interviews mit Nutzern der Community of Practice Plattform

Im Rahmen der Analyse der Community of Practice Plattform wurden nach fünfmonatiger Nutzungszeit im Juni 2016 Interviews mit 17 Personen in Slowenien durchgeführt. Es wurden absichtlich Personen mit unterschiedlicher Nutzungsintensität ausgewählt, um Informationen von Nicht-Nutzern einholen zu können. Während der Interviews wurde ein Forscher aus dem Projekt EmployID zusammen mit einem neutralen Übersetzer, sprich jemandem, der nicht in der Arbeitsagentur angestellt ist, begleitet. Die Interviews hatten eine Dauer von 30 und 45 Minuten und sollten überwiegend analysieren, inwieweit die Interviewpartner die Community of Practice Plattform zum Austausch und zur Reflexion nutzen.

Im Rahmen dieser Interviews wurden Fragen gestellt, ob die Teilnehmer einerseits die Prompts und andererseits die Möglichkeit selbst Fragen zu definieren gesehen haben (siehe Abbildung 4-2 (2/5) und (3)). Die Fragen zielten darauf ab, Feedback von den Nutzern zu bekommen, ob die Prompts in der implementierten Form in der täglichen Praxis hilfreich sind, oder ob diese bspw. störend wirken können. Die Fragen wurden

allerdings nicht in allen Interviews gestellt, da einerseits die Nutzer teilweise direkt gesagt haben, dass sie die Plattform nicht nutzen und andererseits aufgrund von zeitlichen Restriktionen.

Eine inhaltliche Analyse der Interviews hat ergeben, dass nur ein Nutzer sich explizit daran erinnern konnte die Funktionalitäten wahrgenommen und genutzt zu haben. Neun weitere Nutzer haben explizit angegeben, die Funktionalität nicht wahrgenommen zu haben. Den meisten Nutzern ist die Funktion nicht aufgefallen. Einzelne haben gesagt, dass sie überwiegend auf der Plattform gelesen haben und nicht kommentiert haben, was zu der Statistik in Tabelle 4-3 passt. Da die Prompts in der Nähe der Textbox für Beiträge platziert sind, sind sie in den Fällen nicht im Fokus der Nutzer.

Inhaltlich lässt sich dies so interpretieren, dass die Funktionalitäten vom Design her nicht aufdringlich waren und nicht als störend wahrgenommen wurden. Vielmehr lässt dies den Schluss zu, dass die Funktionalitäten nicht selbsterklärend genug für den Großteil der Nutzer waren und das weitere Arbeit notwendig ist, um das Design der Prompts zu optimieren. Die Ergebnisse können auch dahingehend interpretiert werden, dass die Sichtbarkeit der Prompts in dieser Analyse überschätzt wird.

#### **4.4.6 Diskussion des Anwendungsfalles in Slowenien**

An dieser Stelle werden die unterschiedlichen Ergebnisse der Analysen bewertet und in den Kontext der Literatur gesetzt. Zunächst soll an dieser Stelle unterstrichen werden, dass es technische Probleme beim Erfassen aller Prompt-bezogenen Daten gab, wodurch die Datenbasis nicht alle Daten abbildet. Dies schwächt die Aussagekraft der Daten.

Der Anteil der aktiven Nutzer (siehe Tabelle 4-3) ist mit zehn Prozent niedriger als oft in der Literatur genannt (Wenger, McDermott und Snyder 2002, 56). Ein Ergebnis der Interviews ist auch, dass die Moderatoren umfangreiche und komplexe Antworten vorgelebt haben und die Nutzer daher das Gefühl hatten, ebenfalls auf diesem Level Antworten produzieren zu müssen. Dadurch hatten mehrere Interviewpartner das Gefühl, dass eine Teilnahme an der Community sehr viel Zeit in Anspruch nimmt. Dies kann ein Grund für eine niedrigere Teilnehmerzahl sein. Die Nutzung der Community of Practice wurde vonseiten des Managements nie offiziell unterstützt (Blunk und Prilla 2017a), wodurch viele Nutzer das System nur sporadisch, teilweise vor Beginn der eigentlichen Arbeitszeit, genutzt haben. Es ist aber festzuhalten, dass sofern ein Beitrag Antworten bekommen hat, die Anzahl der Antworten überdurchschnittlich im Vergleich zu anderen Communities of Practice ist (Jones und Churchill 2009; Kimmerle, Bientzle und Cress 2014).

Aus Tabelle 4-4 und Tabelle 4-5 lässt sich ableiten, dass das Vorhandensein von Prompts das Erstellen von neuen Themen aufseiten der Nutzer minimal begünstigen könnte. Hier ist der Unterschied aber sehr klein. Bei Antworten scheint dies gegenteilig zu sein, und dies impliziert, dass Prompts das Erstellen von Antworten behindern könnten. Hier sind die Unterschiede zwar minimal, aber leicht höher als bei neuen Themen. Dieser geringe Unterschied durch Prompts ist in anderen Studien aufgefallen (bspw. Renner u. a. 2016). Da mehrere Nutzer in den Interviews ausgesagt haben, dass sie die Prompts und die Möglichkeit eigene Fragen zu stellen nicht wahrgenommen haben, stellt sich die Frage, wie ausgeprägt der Effekt wirklich war. Es lässt sich aufgrund der beiden Tabellen sagen, dass Prompts keine Wirkung auf die Häufigkeit von Beiträgen haben, was bedeutet, dass sie in diesem Anwendungsfall keine neuen Beiträge begünstigen.

Eine Ausnahme können die nutzergenerierten Prompts sein, die auftauchen, wenn jemand eigene Fragen an einen Beitrag anhängt (siehe Abbildung 4-2 (3)). Aus Tabelle 4-7 geht hervor, dass nutzergenerierte Prompts häufiger in Verbindung mit Antworten auftreten. Dies wird durch einige wenige nutzergenerierte Prompts verursacht, da die Mehrheit der nutzergenerierten Prompts (siehe Tabelle 4-6) keine Antworten nach sich gezogen haben. Eine inhaltliche Analyse der Texte zeigt jedoch, dass es eine Korrelation zwischen nutzergenerierten Prompts und Texten gibt, die erfahrungsbasierte Lösungen beinhalten (siehe Tabelle 4-10). Hier ist zu beachten, dass die Anzahl der Fälle in der Korrelationsanalyse mit  $n=30$  vergleichsweise niedrig ist. Diese Ergebnisse implizieren jedoch, dass nutzergenerierte Prompts einen interessanten Effekt in kollaborativer Reflexion haben könnten, der in weiteren Studien untersucht werden sollte. Neue Ideen, wie nutzergenerierte Prompts gestaltet werden können, sind in Abschnitt 7.2.

Der auf Lernen ausgerichtete Prompt (*“What have you learned from this discussion so far?”*) für die Ziele single loop (Z9) und double loop learning (Z10) wurde Nutzern mit Abstand am häufigsten gezeigt und taucht zusammen mit der höchsten Anzahl von Antworten auf (siehe Tabelle 4-6). Ein Grund für die Häufigkeit kann sein, dass der Prompt in der Heuristik, in welcher Situation dieser Prompt angezeigt werden soll, sehr oft vertreten war (Autor, Antwortender, oder Leser; siehe Tabelle 4-1). Die Anzahl der Antworten, die aus diesem Prompt resultieren, ist ähnlich der Anzahl der Antworten bei den Prompts, die beim Schreiben von Antworten angezeigt werden. Durch die erhöhte Zahl der Beiträge lässt sich bei diesem Prompt jedoch am ehesten eine inhaltliche Analyse durchführen (siehe Tabelle 4-11). Diese zeigt, dass der lernbezogene Prompt oft in Verbindung mit Texten steht, in denen double loop learning kodiert ist (siehe Abschnitt 3.4). Dies impliziert, dass der Prompt in der alltäglichen Praxis in einer Arbeitsagentur Lernen in einer Diskussion hervorrufen könnte.

Es hat sich gezeigt, dass die Funktionalität Prompts auf der Startseite anzuzeigen, um die Mitglieder der Community dort zum Schreiben eines Beitrags zu motivieren, nicht funktioniert hat (siehe Tabelle 4-7). Dieses Element muss daher grundlegend neugestaltet werden.

Die Analyse des Lern-Prompts zeigt sehr deutlich, dass eine deutlich größere Datenbasis notwendig ist, um die Wirkung von Prompts in der alltäglichen Praxis zeigen zu können. Dementsprechend kann eine Analyse der übrigen Ziele nicht so durchgeführt werden, als dass eine aussagekräftige Beurteilung möglich ist. Dennoch zeigt die inhaltliche Kodierung der Texte, dass 46 % der Texte mit den Prompts übereinstimmen, was eine Wirkung von Prompts impliziert (siehe Tabelle 4-8).

Reflexion wird oft ausgelöst, wenn sich eine Situation anders verhält, als man es antizipiert hat, bspw., wenn man in einer Situation überrascht wird (siehe Abschnitt 2.4). Interessanterweise traten die beiden Prompts, die für dieses Ziel entwickelt sind (Z1), nie in Zusammenhang mit Antworten auf („*Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?*“ und „*What was unexpected for you in your daily work in the last 2 weeks? Why?*“; siehe Tabelle 4-6). Daraus lässt sich schließen, dass zwar in der täglichen Praxis die Überraschung ein passender Auslöser für Reflexion ist, aber eine gezielte Frage nach solchen Überraschungen nicht zwingend den Wunsch auslöst über ein bestimmtes Thema schriftlich in einer Community of Practice mit anderen zu reflektieren. Dies kann auch kulturelle Gründe haben, wie bspw., dass ungern Fehler oder fehlende Kenntnisse offenbart werden. Es ist möglich, dass die Empfänger darüber gedanklich reflektiert haben, was sich jedoch anhand der Daten nicht belegen lässt. Dementsprechend sind weitere Studien notwendig, um einen passenden Startpunkt für schriftliche kollaborative Reflexion zu finden.

#### **4.5 Anwendungsfall: Kroatien**

Die in der Arbeitsagentur in Slowenien durchgeführte Studie wurde ebenfalls in der Arbeitsagentur in Kroatien durchgeführt. Da sich die Struktur der Arbeitsagentur in Kroatien von der in Slowenien unterscheidet (Blunk und Prilla 2017a), wurden in diesem Anwendungsfall die gleichen Prompts genutzt, um eine breitere Datenbasis für die Evaluation des Konzepts zu erhalten.

Da die beiden Unternehmen unterschiedlicher Größe und interner Organisation sind, unterschiedliche Abläufe haben und die Studie einen längeren Zeitraum je Kondition hatte, wurden die Studien getrennt analysiert.



#### 4.5.1 Kontext des Anwendungsfalles in Kroatien

Es ist zu beachten, dass ursprünglich zwei Gruppen in einem within-subjects Design geplant waren, was aufgrund der niedrigen Zahl der anfänglichen Nutzer und der sich konstant verändernden Umgebung anders ausgewertet wurde. Es wurden neuregistrierte Nutzer bis zum Ende in das Experiment mit aufgenommen. Jeder Nutzer hat einen eigenen persönlichen Zeitraum, der festlegt, wann der Nutzer Prompts sieht und wann keine Prompts angezeigt werden. Nutzer, deren Registrierungszeitpunkt älter als die Dauer der Studie ist, haben *keine* Prompts angezeigt bekommen. Bis zum Ende des Zeitraums, in dem Daten gesammelt wurden, wurden neu registrierte Benutzer in die Studie mit aufgenommen. Dies ist analog zu der Studie im Anwendungsfall in Slowenien (siehe Abschnitt 4.4.1).

Die Tabelle 4-12 zeigt den zeitlichen Ablauf der Studie. Die Studie ist im Oktober 2016 mit einem initialen Workshop gestartet, in dem unterschiedlichen Beratern die Community of Practice Plattform vorgestellt wurde. Diese Personen haben angefangen auf der Plattform zu diskutieren und ihre Kollegen eingeladen. In Kroatien wurden anschließend zusammen mit den Moderatoren unterschiedliche Gruppen von Personen identifiziert, die von der Plattform profitieren können und diese wurden sukzessive eingeladen. In der Tabelle stehen Zeitpunkte, wann die Nutzer in die nächste Gruppe innerhalb der Studie wechseln. Jeder Benutzer ist drei Monate in jeder Kondition, was bedeutet, dass er drei Monate lang Prompts sieht und danach drei Monate keine Prompts mehr sieht (oder vice versa). Sobald sich ein Nutzer registriert und eingeloggt ist, beginnt der Studienzeitraum für diesen Nutzer. Nach Ablauf der Studie haben Nutzer im Unterschied zum Anwendungsfall in Slowenien *keine* Prompts angezeigt bekommen. Der Zeitraum wurde in diesem Anwendungsfall, im Vergleich zum Anwendungsfall in Slowenien länger gestaltet, um eine höhere Chance auf Sammeln einer größeren Datenmenge zu haben.

**Tabelle 4-12: Ablauf der Studie in Kroatien**

Datum	Event
04.10.2016	Start der Community in einem Workshop mit 25 Personen.
05.01.2017	Nutzer im within-subjects Design wechseln die Kondition (von Prompt zu Nicht-Prompt und von Nicht-Prompt zu Prompt).
05.04.2017	Für die Nutzer, die seit dem initialen Workshop registriert sind, endet die ursprüngliche Laufzeit der Studie.
28.02.2018	Letzter Tag, an dem Nutzungsdaten gesammelt wurden, bevor mit der Analyse für den Projektabschluss begonnen wurde.

Eine Auflistung, welche Faktoren die Entwicklung der Community begünstigt oder behindert haben, findet sich in (Blunk und Prilla 2017a). Eine Darstellung, wie sich die

Community hinsichtlich der Zahl der Mitglieder und der Themengruppen entwickelt hat, findet sich in (EmployID 2018, 104–109).

Für das allgemeine Vorgehen in der Analyse siehe Abschnitt 4.3.3. Im Folgenden werden nur spezifische Analyseschritte für diesen Anwendungsfall samt den Ergebnissen beschrieben.

#### 4.5.2 Deskriptive Analyse des Anwendungsfalles in Kroatien

An dieser Stelle wird eine Übersicht über die Community selbst gegeben. Es haben sich bis zum Analysezeitpunkt 155 Personen registriert (Tabelle 4-13; von 1600 Mitarbeitern), wodurch diese Community im Vergleich zum Anwendungsfall in Slowenien (siehe Abschnitt 4.4) deutlich kleiner ist.

Ungefähr ein Drittel der registrierten Mitglieder ist als aktiv zu bezeichnen und hat zumindest einen Beitrag geschrieben und Beiträge im Forum gelesen. Von den Nutzern ist ein Nutzer mit 20 Beiträgen (Themen und Antworten) deutlich aktiver als die übrigen Nutzer (die Person mit den zweitmeisten Beiträgen hat 14 Beiträge verfasst). Der Anteil der inaktiven Personen liegt bei ca. 15 %. Es handelt es sich um Personen, die sich registriert haben und die einzelnen Bereiche der Plattform betrachtet haben, aber den eigentlichen Bereich der Community, das Forum, absichtlich nicht besucht oder nicht gefunden haben und somit keine Beiträge gelesen haben. Im Durchschnitt hat ein Beitrag in dem Forum ca. vier Antworten bekommen, was vergleichbar mit anderen Communities ist, wo je nach Forum 3,2 bis 4,6 (Jones und Churchill 2009; Kimmerle, Bientzle und Cress 2014) Antworten je Thema geschrieben wurden.

**Tabelle 4-13: Gesamtaktivität der Benutzer und Moderatoren in der Community  
in Kroatien zwischen 10.2016 und 02.2018**

<b>Aktivität</b>	<b>Nutzer</b>	<b>Moderatoren</b>
Anzahl registrierte Personen	155	4
Anzahl aktive Personen (lesen und schreiben)	47	/
Anzahl passive Personen (nur lesen)	85	/
Anzahl inaktive Personen (nur registriert)	23	/
Anzahl Lese-Events	2872	496
Anzahl geschriebene Themen	27	13
Anzahl geschriebene Antworten	134	29

Durch die hohe Zahl der Antworten von Nutzern und die deutliche niedrigere Zahl der Antworten von Moderatoren lässt sich entnehmen, dass die Nutzer untereinander viel diskutieren und die Moderatoren eher neue Themen beisteuern.

### 4.5.3 Quantitative Analyse des Anwendungsfalles in Kroatien

Die Idee war ein within-subjects Design, um bei kleineren Nutzermengen den Vorteil zu haben, dass jeder Nutzer während der Nutzung Prompts angezeigt bekommt.

Eine Analyse der Daten ergibt, dass es einen Nutzer gibt (anonymisierte Nutzer-ID u/18137), der in einer Kondition deutlich aktiver als die übrigen Nutzer war (bspw. 20 verfasste Beiträge (M=3,42; SD=3,7; Mdn.=2)). Diese Person hat einerseits sehr viele Beiträge verfasst und hat andererseits den Großteil der Beiträge in einer der beiden Konditionen *prompt* bzw. *Nicht-prompt* verfasst. Anschließend war die Person nur wenig aktiv. Daher ist unklar, ob die Unterschiede zwischen den Konditionen bei den Prompts liegt oder andere Einflüsse hat. Daher wurde der betreffende Nutzer ausgeschlossen.

Da in dieser Studie nach Ablauf des Studienzeitraums bei jedem Nutzer *keine* Prompts mehr angezeigt wurden, haben die meisten Nutzer eine wesentlich längere Zeit ohne Prompts als mit Prompts auf der Plattform verbracht (Zeitraum der Datenerfassung: 10/2016-02/2018). Es folgt eine Gegenüberstellung der Aktivität zwischen der Nicht-Prompt-Kondition und der Prompt-Kondition einerseits in dem ursprünglichen Studienzeitraum (drei Monate je Kondition), und einmal über den kompletten Datenerfassungszeitraum. Lese-Events auf Foren-Ebene sind Lesezugriffe auf die Liste aller Beiträge in einem Forum (siehe Abbildung 4-2 (1)) und Lese-Events auf Beitragsebene bedeutet, dass ein Nutzer die Beiträge innerhalb einer Diskussion gelesen hat (siehe Abbildung 4-2 (4)). In den letzten beiden Zeilen findet sich eine Quote der neuen Beiträge anteilig an den jeweiligen Schreibzugriffen, um die Zahlen zu normalisieren.

**Tabelle 4-14: Lese- und Schreibraten der Nutzer (ohne Nutzer mit anonymisierter ID u/18137) und Moderatoren in den beiden Konditionen der Studie (nur Zeiträume der Konditionen)**

	Nutzer		Moderatoren	
	Keine Prompts	Prompts	Keine Prompts	Prompts
Alle Lese-Events	1660	1476	185	212
Lese-Events Forum	470	484	28	49
Lese-Events Beiträge	1190	992	157	163
Erstellte Themen	5	8	3	3
Erstellte Antworten	59	45	5	21
Lese-Schreib-Quote Forum	1,06 %	1,65 %	10,71 %	6,12 %
Lese-Schreib-Quote Beitrag	4,96 %	4,54 %	3,18 %	12,88 %

Tabelle 4-14 betrachtet den ursprünglichen Zeitraum der Studie von zwei mal drei Monaten je Mitglied auf der Plattform. Es ist festzustellen, dass sich bei Nutzern beim Vergleich der beiden Konditionen die Werte jeweils um ein halbes Prozent unterscheiden.

Wenn Nutzer Prompts angezeigt bekommen werden relativ betrachtet bei mehr neuen Themen pro Seitenaufruf weniger neue Antworten pro Seitenaufruf geschrieben. Die Aktivität ist in beiden Konditionen ähnlich groß, obwohl in der Kategorie ohne Prompts mehr Antworten geschrieben wurden. Der Effekt der Prompts ist also vernachlässigbar. Prompts könnten sich eher eignen, um neue Themen zu ermutigen und weniger dazu um neue Antworten zu ermutigen. Diese erste Analyse betrachtet ausschließlich die Häufigkeit der Antworten, aber nicht die Qualität der Antworten.

Die Werte der Moderatoren sind der Vollständigkeit halber aufgeführt. Allerdings ist dort aufgrund starker Schwankung in der Zahl der Beiträge bzw. Lesezugriffe je Kondition der relative Vergleichswert wenig aussagekräftig. Insgesamt haben die vier Moderatoren im Vergleich zu den Nutzern relativ mehr Aktivität als die Nutzer gezeigt. Ein Grund dafür ist, dass die Moderatoren die Aufgabe haben, die Mitglieder in der Community of Practice zum Partizipieren bewegen sollen und dies häufig über eigene Beiträge geschieht (Wenger, McDermott und Snyder 2002). Zusätzlich wurden die Moderatoren ausgewählt, weil sie initial eine sehr hohe Motivation zur Gründung einer Community hatten und das Thema innerhalb der Organisation vorantreiben wollten. Somit begünstigen die Faktoren eine hohe intrinsische Motivation zur Arbeit in der Community und die Aufgabe als Moderator die Aktivität in der Community zu fördern. Inwieweit die Prompts einen Unterschied gemacht haben, lässt sich aus diesen Daten nicht analysieren.

Nachdem Tabelle 4-14 den ursprünglichen Zeitraum der Studie von zwei mal drei Monaten je Benutzer beleuchtet hat, zeigt Tabelle 4-15 die Aktivität über den gesamten Datenerfassungszeitraum je Nutzer, sprich von 10/2016 bis 02/2018. Wenn sich ein Nutzer direkt zu Beginn registriert hat, nimmt er erst an der Studie teil und sieht anschließend *keine* Prompts mehr. Auf den ersten Blick lässt sich feststellen, dass sich die Raten bei Nutzern umgekehrt haben. Jetzt scheint die Erstellung von Antworten zu begünstigen und die Erstellung von neuen Themen zu behindern. Dieses Ergebnis impliziert, dass in diesem System zu Zeiten, wo den Nutzern Prompts über längere Zeiträume gezeigt werden, dies eher die Erstellung von neuen Themen beeinflusst. Zusätzlich scheint das geringere Anzeigen von Prompts eher das Erstellen von Antworten zu beeinflussen. Auf der Seite der Moderatoren ist das Verhältnis und größtenteils gleichgeblieben.

Tabelle 4-15 impliziert, dass während die Prompts gezeigt wurden, weniger Themen erstellt wurden. Auf der anderen Seite scheinen die Prompts zu etwas mehr Antworten zu führen.

Aus den quantitativen Daten lässt sich kein eindeutiger Grund für das Umkehren der Quoten aufseiten der Nutzer ableiten. Es ist zu vermuten, dass die Prompts keinen ausschlaggebenden Einfluss auf die Häufigkeit haben, in der jemand ein neues Thema erstellt oder eine Antwort verfasst.

**Tabelle 4-15: Lese- und Schreibraten der Nutzer (Nutzer ohne mit anonymisierter ID u/18137) und Moderatoren in den beiden Konditionen der Studie (kompletter Datenerfassungszeitraum)**

	Nutzer		Moderatoren	
	Keine Prompts	Prompts	Keine Prompts	Prompts
Alle Lese-Events	2497	1476	412	212
Lese-Events Forum	747	484	94	49
Lese-Events Beiträge	1750	992	318	163
Erstellte Themen	19	8	10	3
Erstellte Antworten	68	45	8	21
Lese-Schreib-Quote Forum	2,54 %	1,65 %	10,64 %	6,12 %
Lese-Schreib-Quote Beitrag	3,89 %	4,54 %	2,52 %	12,88 %

Prompts werden nur als gelesen gewertet, wenn diese auf dem Bildschirm im Sichtfeld des Nutzers waren (siehe Abschnitt 4.3.3). Laut Tabelle 4-16 wurden in Kroatien im gesamten Studienzeitraum 2577 Prompts vom Server abgerufen, die das System zu ca. 68 % als im Sichtfeld des Nutzers klassifiziert hat. Während diese Prompts für Nutzer sichtbar waren, sind 85 Beiträge entstanden, was einer Quote von ca. fünf Prozent entspricht. In der Tabelle 4-16 ist konkret aufgeschlüsselt, welcher Prompt, wie oft gesehen wurde und wie oft ein Beitrag entstanden ist. Die Anzahl, wie oft ein Prompt sichtbar war, ist stark unterschiedlich. Dies liegt an der Konfiguration der Prompts bzw. der Ziele (siehe Abschnitt 3.2), nach der bestimmte Ziele bspw. die Dokumentation von Gelerntem seltener auftreten als andere Ziele wie bspw. die Dokumentation eigener Erfahrungen.

In Tabelle 4-16 sind mehrere Prompts aufgeführt, die zwar gesehen wurden, aber bei denen nie ein Beitrag erfolgt ist („*Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?*“ (Z1) und „*What was unexpected for you in your daily work in the last 2 weeks? Why?*“ (Z1)). Hier liegt die Vermutung nahe, dass diese in Zukunft ausgeschlossen werden können, zumal diese jeweils ca. 250x im Sichtfeld des Nutzers waren. Hier liegt die Interpretation nahe, dass etwas Unerwartetes ein Auslöser für Reflexion ist, während man sich selbst innerhalb der Situation befindet (siehe Abschnitt 2.4.2; Schön 1983). Dies scheint aber nicht über Prompts realisierbar und daher können diese Prompts im Praxiseinsatz vernachlässigt werden.

Die übrigen Prompts im Bereich zur Anregung neuer Themen sind von der Häufigkeit her ähnlich oft zusammen mit Prompts aufgetaucht. Diese sollten in weiteren Studien beibehalten werden.

Bei den Prompts, die innerhalb einer Diskussion, also beim Schreiben von Antworten, gezeigt wurden, gibt es bei näherer Betrachtung weniger Unterschiede. Der Prompt, der am meisten gezeigt wurde („*What have you learned from this discussion so far?*“; Z9, Z10) tauchte zusammen mit den meisten Beiträgen (28) auf. Die Häufigkeit von Beiträgen, die bei diesem Prompts erfolgte, ist vergleichbar mit den anderen Prompts. Bei zwei Prompts wird anteilig zur Frequenz, wie oft der Prompt gesehen wurde, häufig ein Beitrag verfasst („*Do you already have an idea how you want to tackle similar situations in future? If not, how about phrasing a new question?*“ (Z11) und „*What specific steps do you want to take to approach similar situations in future?*“ (Z12)). Da der Prompt aber nur sehr selten gezeigt wurde, ist die Quote höher. Er wurde seltener angezeigt, weil die meisten Themen erst eine notwendige Anzahl an Antworten erhalten musste, bevor neue Prompts für den Autor angezeigt wurden (siehe Abschnitt 4.3.2). Drei Prompts in dieser Kategorie haben explizit den Autor eines Beitrags als Ziel, damit dieser aus der Diskussion heraus Lösungswissen in Betracht zieht und konkrete Schritte für die Zukunft plant (siehe Kapitel 3 und siehe Tabelle 4-1; die beiden vorherigen Prompts und der folgende: „*Have you already experiences after making changes in your approach? If yes, what are those experiences?*“ (Z13)). Aufgrund der niedrigen Zahlen, wie oft diese Prompts vom Server abgerufen wurden, lässt sich schließen, dass dieser Fall einerseits nicht sehr oft eingetreten ist und andererseits, dass die Autoren nicht sehr oft die eigenen Beiträge besuchen, um ggf. später darüber zu reflektieren. Die Empfehlung ist an dieser Stelle alle Prompts beizubehalten, da sich keiner weder besonders hervorgehoben hat noch als wirkungslos identifiziert wurde. Insbesondere bei den Prompts, die den Autor eines Themas zum Planen konkreter Schritte für die Zukunft leiten sollen, fehlen Daten für eine konkrete Empfehlung.

Der nächste Bereich sind Prompts, die für andere gezeigt werden, falls der Autor eines Themas diese bei Erstellung des Themas ausgewählt hat (siehe Abbildung 4-2 (3)). Es lässt sich feststellen, dass die vordefinierten Fragen #1 und #3 nie ausgewählt wurden. Die vordefinierte Frage #2 wurde für insgesamt drei Themen ausgewählt und befand sich in der Hälfte der Seitenaufrufe im Sichtfeld des Nutzers. Ein Autor hat bei Erstellung seines Themas zwei Fragen ausgewählt (#2 und #3), welche in dem Thema bei drei Antworten präsent waren. Hier ist die Sortierung einfach, da Prompt #1 nie genutzt wurde und dementsprechend in Zukunft nicht weiter genutzt werden sollten. Die Prompts sollten in künftigen Studien näher analysiert werden.

Insgesamt drei Nutzer haben die Funktionalität genutzt und eigene Beiträge geschrieben. Bei einem der Beiträge sind Antworten in Anwesenheit von Prompts verfasst worden. Die anderen beiden Themen haben insgesamt zwei Antworten bekommen, aber nicht während ein Nutzer oder Moderator in der Prompt-Kondition war.

Aufgrund der Definition der einzelnen Ziele bzw. der Heuristik für die Anzeige bestimmter Prompts, zeigt Tabelle 4-16 stark unterschiedliche Häufigkeiten, wie oft ein Prompt angezeigt wurde. Dies liegt daran, dass bspw. innerhalb einer Diskussion die Prompts, welche den Autor eines Themas bitten aus den Antworten einen Plan für die Zukunft abzuleiten, nur angezeigt werden, sobald Antworten vorhanden sind. Zusätzlich werden diese Prompts aufgrund der Formulierung des Prompts auch nur dem Autor eines Themas angezeigt. Andere Prompts, wie bspw. die Aufforderung an Leser eigene Erfahrungen beizusteuern, haben diese Voraussetzungen nicht und werden dementsprechend häufiger angezeigt. Zusätzlich wurden durch die Konfiguration der Ziele und der Voraussetzungen im System manche Prompts häufiger ausgewählt als andere.

**Tabelle 4-16: Übersicht welcher Prompt in Kroatien wie oft gesehen wurde und wie oft ein Beitrag geschrieben wurde (gruppiert nach den Orten: Prompts bei Erstellung neuer Themen, bei der Erstellung neuer Antworten und auf der Startseite (siehe auch Abbildung 4-2 und Abbildung 4-3 auf S. 127))**

<b>Ziele</b>	<b>Prompt-Text</b>	<b># Prompt dargestellt</b>	<b>% Prompts im Sichtfeld</b>	<b>#Beiträge</b>	<b>% Beiträge bei sichtbaren Prompts</b>	<b>Ort</b>
Z1	Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?	426	57,28 %	0	0 %	Neue Disk.
Z1	What was unexpected for you in your daily work in the last 2 weeks? Why?	420	59,76 %	0	0 %	Neue Disk.
Z4	Try to describe your point of view using a metaphor.	84	63,10 %	4	8 %	Neue Disk.
Z4	What do you think about what has happened? Do you already know how you want to approach this differently in future?	84	66,67 %	2	4 %	Neue Disk.
Z2	Try to make clear what you want to know from your colleagues	171	57,89 %	7	7 %	Neue Disk.
Z5	What did you already try to approach the situation?	177	66,10 %	8	7 %	Neue Disk.
Z8	What would you personally suggest from your experience? Why?	174	74,14 %	8	6 %	Innerhalb Disk.
Z6	Do you agree / disagree with any of the existing comments? What would you suggest from your experience? Why?	180	77,78 %	8	6 %	Innerhalb Disk.
Z9, Z10	What have you learned from this discussion so far?	521	82,34 %	28	7 %	Innerhalb Disk.
Z11	Do you already have an idea how you want to tackle similar situations in future? If not, how about phrasing a new question?	17	82,35 %	2	14 %	Innerhalb Disk.
Z12	What specific steps do you want to take to approach similar situations in future?	2	100,00 %	1	50 %	Innerhalb Disk.



<b>Ziele</b>	<b>Prompt-Text</b>	<b># Prompt dargestellt</b>	<b>% Prompts im Sichtfeld</b>	<b>#Beiträge</b>	<b>% Beiträge bei sichtbaren Prompts</b>	<b>Ort</b>
Z13	Have you already experiences after making changes in your approach? If yes, what are those experiences?	15	80,00 %	1	8 %	Innerhalb Disk.
Z7	What do you think about how I dealt with this situation?	0	0,00 %	0	0,00 %	QS-Plugin #1
Z8	What would you personally from your experience recommend as a solution to the case and why?	98	51,02 %	10	20,00 %	QS-Plugin #2
Z6	Have you ever had a similar experience? What happened?	0	0,00 %	0	0,00 %	QS-Plugin #3
Z6, Z8	What would you personally from your experience recommend as a solution to the case and why? Have you ever had a similar experience? What happened?	58	63,79 %	2	5,41 %	QS-Plugin #2 + #3
/	What would you personally from your experience recommend as a solution to the case and why? What interest is placed as a priority and what are the closest occupations related to the completed school? <sup>23</sup>	11	81,82 %	0	0,00 %	Nutzer + Vordefiniert
/	What do you think would be the best solution? One day do not work with clients like CZS? Do not agree that there are no competitions for 1-2 workers in the season? Increase the number of counselors? <sup>23</sup>	33	96,97 %	0	0,00 %	Nutzer
/	Can you please suggest the items from the house rules that you use, since we do not have the current impression of what we might be able to state, except for the obligatory preservation of office equipment of course. <sup>23</sup>	45	73,33 %	4	12,12 %	Nutzer
/	What are the benefits and challenges, according to your experience? <sup>23</sup>	30	96,67 %	0	0,00 %	Nutzer
Gesamt Abrufe/Beiträge und Durchschnitt gesehen/Beitrag verfasst		2546	68,19 %	85	4,9 %	

<sup>23</sup> Übersetzung aus dem Kroatischen mit Google Translate (<https://translate.google.com/>; letzter Zugriff am 12.08.2019)

**Tabelle 4-17: Sichtbarkeit pro Prompt-Kategorie und Ort in Kroatien**

<b>Ort</b>	<b>Nutzer</b>			<b>Moderatoren</b>		
	Anteil im Log	Anteil wie oft sichtbar	Anteil wie oft Beitrag geschrieben von sichtbar	Anteil im Log	Anteil wie oft sichtbar	Anteil wie oft Beitrag geschrieben von sichtbar
In der Liste der Themen: Prompt vom Server abgerufen	41,10 %	41,10 %		12,00 %	12 %	
In der Liste der Themen: Prompt im Sichtfeld	56,26 %	56,26 %		82,00 %	88 %	
Ein neues Thema wurde erstellt	2,64 %		4,48 %	6,00 %		6,82 %
Während des Lesens eines Beitrags: Prompt vom Server abgerufen	20,72 %	20,72 %		12,10 %	12,10 %	
Während des Lesens eines Beitrags: Prompt im Sichtfeld	74,32 %	79,28 %		76,61 %	87,9 %	
Eine neue Antwort wurde erstellt	4,96 %		6,26 %	11,29 %		12,84 %
Startseite: Prompt vom Server abgerufen	41,97 %			40,80 %		
Startseite: Prompt im Sichtfeld	58,03 %			59,20 %		
Vom Nutzer überschriebener Prompt vom Server abgerufen	32,20 %	32,2 %		21,05 %	21,05 %	
Vom Nutzer überschriebener Prompt im Sichtfeld	63,56 %	67,8 %		63,16 %	78,95 %	
Eine neue Antwort wurde erstellt	4,24 %		6,25 %	15,79 %		

Die Tabelle 4-17 zeigt, wie oft ein Prompt, der für einen bestimmten Ort vom Server abgerufen wurde, im Sichtfeld war und bei der Verfassung eines Themas bzw. einer Antwort präsent war. Es lässt sich feststellen, dass bei Nutzern ein Prompt bei der Erstellung eines neuen Themas (siehe Abbildung 4-2 (2)) nur in ca. 56 % der Fälle sichtbar war. Dies liegt daran, dass der Prompt dort platziert ist, wo später die Aktion stattfindet (McKenzie-Mohr 2000), sprich bei der Textbox, in der das neue Thema verfasst wird). Hier wurde in ca. 4,5 % der Fälle ein Beitrag verfasst. Bei Moderatoren ist sowohl der Anteil, wie oft ein Prompt in der Sektion sichtbar war als auch der Anteil der geschriebenen Beiträge höher. Eine Interpretation ist, dass Moderatoren allein aufgrund ihrer Rolle und Motivation mehr Beiträge verfasst haben und daher häufiger bis an das Ende der Liste der Themen navigiert haben, um einen neuen Beitrag zu verfassen. Die normale Nutzung eines Forums beinhaltet, dass man sich die neuen Themen anschaut bzw. die Themen, in denen zuletzt eine neue Antwort verfasst wurde. Das Forum ist so konfiguriert, dass die neuesten Themen bzw. die Themen, die zuletzt eine Antwort bekommen haben, oben stehen. Vermutlich navigiert ein normaler Nutzer nicht bei jedem Besuch der Seite die lange Liste der Themen nach unten, um den Prompt zur Verfassung eines neuen Themas zu sehen. Wenn allerdings neue Gruppen gegründet werden, dann sind nicht viele Themen vorhanden und Nutzer finden den Editor zur Erstellung neuer Themen eher. Ein Ergebnis der Auswertung ist, dass die Möglichkeit ein neues Thema zu verfassen prominenter dargestellt werden muss.

Innerhalb eines Themas wurden Prompts bei Nutzern und Moderatoren ähnlich oft gesehen (in ca. 79 % bis 88 % der Fälle). Hier sind ebenfalls bei Moderatoren mehr Antworten entstanden als bei Nutzern. Die Interpretation ist hier, dass Moderatoren aufgrund ihrer Rolle eher helfend zur Seite stehen wollten. Die Moderatoren haben in Summe nicht sehr viele Antworten verfasst (siehe Tabelle 4-13 oder Tabelle 4-15). Es ist hervorzuheben, dass in dieser Analyseperspektive, Prompts häufiger beim Verfassen neuer Antworten eine Rolle spielen.

Aus der Tabelle 4-17 lässt sich entnehmen, dass aufgrund des Prompts auf der Startseite keine neuen Themen verfasst wurden. Dementsprechend ist das Element als wirkungslos zu betrachten. Im Durchschnitt waren die Prompts auf der Startseite ähnlich oft zu sehen, wie die Prompts für neue Themen. Eine Interpretation für den ausbleibenden Erfolg ist, dass die Nutzer die Startseite als Übergang in den Forenbereich genutzt haben und dort in erster Linie gelesen haben. Womöglich haben die Nutzer eher einen Beitrag verfasst, nachdem sie in anderen Beiträgen gelesen hatten, als dass sie direkt die Plattform besucht haben, um einen Beitrag zu verfassen. Eine andere, eher wahrscheinliche Erklärung ist, dass es aus Sicht des mentalen Modells von Nutzern über ein Forum unüblich ist, über die Startseite einen Beitrag verfassen zu wollen. Dieses mentale Modell des Nutzers

könnte die Möglichkeit zur Erstellung neuer Themen ausschließlich in dem Bereich des Forums verorten. Dieses Element bedarf also einer sorgfältigen Neugestaltung.

Nutzergenerierte Prompts, die erstellt wurden, indem über das gesonderte Feature eigene Fragen als Prompts hervorgehoben wurden (siehe Abbildung 4-2 (3)), waren durchschnittlich ca. zehn Prozent weniger im Sichtfeld des Nutzers als reguläre Prompts. Für diese Tatsache gibt es keine logische Erklärung, da für Nutzer beide Arten von Prompts gleich angezeigt werden. Die Anzahl der Antworten, die in Anwesenheit eines nutzergenerierten Prompts geschrieben wurden, ist aber interessanterweise ähnlich hoch wie bei den regulären Antworten (6,26 % und 6,25 %). Dies lässt vermuten, dass die nutzergenerierten Prompts eher akzeptiert werden als die regulären Prompts. Ein Grund dafür kann sein, dass diese besser auf das Thema zugeschnitten sind, weil sie, wie die regulären Prompts, keiner einfachen Heuristik entsprechen, sondern von einem Menschen für die entsprechende Situation entworfen wurden. Es ist möglich, dass die Formulierung passender für die jeweilige Situation bzw. den organisationalen Kontext ist.

#### 4.5.4 Qualitative Analyse des Anwendungsfalls in Kroatien

In diesem Abschnitt wird eine qualitative Perspektive für die Analyse gewählt. Die von den Nutzern und Moderatoren erstellten Beiträge wurden durch zwei Muttersprachler inhaltlich auf Reflexion untersucht und entsprechend kodiert (siehe Abschnitt 4.3.3).

Zusätzlich zu dem bisherigen Kodierschema (siehe Abschnitt 3.4.1), wurden bei dieser Analyse drei Codes einbezogen, die beurteilen sollten, wie sehr der Nutzer in einer Antwort bzw. einem Thema auf einen Prompt Bezug nimmt (siehe Abschnitt 3.4.2).

Das durchschnittliche Krippendorffs-Alpha für die reflexionsbezogenen Codes liegt bei .92 (min. .85 und max. 1). Das durchschnittliche Krippendorffs-Alpha für die drei neuen Prompt-bezogenen Codes liegt bei .88 (min. .86 und max. .89).

**Tabelle 4-18: Häufigkeit des Einflusses eines Prompts auf ein Thema (Kroatien)**

Prompt-Einfluss auf den Beitrag	Häufigkeit	Gruppierung
Kein Prompt bei Erstellung des Beitrags involviert	122	Kein Einfluss
Code: User did NOT respond to prompt	49	
Code: User responded in PARTS to the prompt	14	Einfluss vorhanden
Code: User responded fully to the prompt	18	

Aus Tabelle 4-18 lässt sich ableiten, wie oft gemäß der qualitativen Inhaltsanalyse einer der Prompts Einfluss auf den Text hatte. Da die Häufigkeiten für die Prompt-Codes nicht sehr hoch sind, werden die Codes für die weitere Analyse in eine neue Variable gruppiert. Die neue Variable *Einfluss* wird als boolesche Variable wie folgt gebildet: Es wird

unterschieden, ob gar kein Einfluss vorhanden war (0), bspw., weil entweder kein Prompt sichtbar war, als der Beitrag geschrieben wurde oder weil nach Meinung der Kodierenden der Text keinen Bezug hat. Ein Einfluss ist vorhanden (1), wenn nach Meinung der Kodierenden der Text entweder teilweise oder voll auf den Prompt passt. Somit hat ein Prompt in ca. 40 % der Fälle, in denen er angezeigt wurde, einen Einfluss auf den Text des Benutzers, was sehr deutlich ist. Es kann sein, dass der Autor eines Themas oder einer Antwort bestimmte Aspekte ohnehin schreiben wollte und dadurch unbewusst dem Prompt entsprochen hat. Dies lässt sich anhand der Daten nicht analysieren.

Im Folgenden wird analysiert, ob die Prompts mit Einfluss auf den Text eine Verbindung zu den Codes aus dem reflexionsbezogenem Kodierschema (siehe Abschnitt 3.4.1) haben. Alle Korrelationen sind im Folgenden mit \* für eine signifikante Beziehung auf dem 0,05 Level und \*\* für das 0,01 Level gekennzeichnet.

Die Tabelle 4-19 impliziert, dass wenn der Autor eines Textes Prompts teilweise oder ganz in seinen Text einbezieht, der Beitrag häufig Erfahrung (EXP), Hinweise (ADV) und erfahrungsbasierten Lösungsvorschläge (SUG\_EXP) beinhaltet.

**Tabelle 4-19: Korrelation zwischen den Prompts mit Einfluss auf den Text und den Codes für Reflexion (n=203 Beiträge, bei denen in 32 Beiträgen ein Einfluss kodiert wurde; 107x Code EXP, 12x Code ADV und 67x Code SUG\_EXP kodiert)**

		EXP	ADV	SUG_EXP
<b>Einfluss</b>	Pearson Correlation	,220**	,178*	,271**
	Sig. (2-tailed)	0,002	0,011	0,000

An dieser Stelle soll ein Bezug zu Tabelle 4-16 hergestellt werden, welche zeigt wie oft welcher Prompt gesehen wurde und wie oft dazu ein Text geschrieben wurde, um neben dem Inhalt auch die Häufigkeit als zusätzliche Perspektive in die Analyse einzubeziehen. Laut der Tabelle gibt es sehr viele Beiträge zu diesen Prompts: „*Do you agree / disagree with any of the existing comments? What would you suggest from your experience? Why?*” (Z6) und „*What have you learned from this discussion so far?*” (Z9, Z10). Der erste Prompt zielt einerseits auf eine Diskussion mit Argumenten für oder gegen Vorschläge ab, was laut dem Kodierschema mit den Codes AGR für Zustimmung und DISAGR für Ablehnung markiert wird. Der andere Code zielt auf Lernen ab, was laut Kodierschemas mit den Codes S\_LOOP und D\_LOOP für Single und Double Loop Learning steht. Keiner der Beiträge, die mit den beiden Prompts in Verbindung stehen, beinhalten einen dieser vier Codes. Dies bedeutet, dass die Prompts zwar in Zusammenhang mit Antworten stehen, aber dass diese Antworten nicht die gewünschten Inhalte haben. Dementsprechend scheinen diese Prompts in diesem Anwendungsfall keine direkte Wirkung auf die Texte zu haben.

Eine Analyse, welcher Prompt, in welchem Umfang Einfluss auf den Text genommen haben könnte, zeigt Tabelle 4-22. Es gibt nur einen Prompt, der 32x in Zusammenhang mit einem Text aufgetaucht ist. Dementsprechend lassen sich diese Daten nicht quantitativ analysieren und bedürfen einer Analyse in künftigen Studien. Die Summe der bewerteten Prompts in Tabelle 4-22 ist mit 81 niedriger als die Anzahl (85) der Prompts in Zusammenhang mit einem Text in Tabelle 4-16. Dies liegt daran, dass die Kodierenden nicht bei allen Prompt-Codes übereingestimmt haben. Die Dateien ohne Übereinstimmung wurden bei der Analyse nicht berücksichtigt.

Es lässt sich feststellen, dass aus den Prompts, die neue Themen beeinflussen sollen (siehe Tabelle 4-16 Z1 bis Z5), nur ein Prompt deutlichen Einfluss auf den Text genommen hat: „Try to make clear what you want to know from your colleagues“ (Z2). Hier muss festgestellt werden, dass dieser Prompt generisch ist, aber durch das Abzielen auf aussagekräftigere Beiträge den reflexiven Austausch ermöglichen soll. Im Bereich der Antworten lässt sich feststellen, dass der Prompt, der nach Lernen fragt („What have you learned from this discussion so far?“ (Z9, Z10)), nur einmal Einfluss auf den Text hatte. Somit ist dieser Prompt in diesem Anwendungsfall nicht als wirksam zu betrachten. Die zwei Prompts, die darauf abzielen, dass konkrete Schritte für die Zukunft geplant werden („Do you already have an idea how you want to tackle similar situations in future? If not, how about phrasing a new question?“ (Z11) und „What specific steps do you want to take to approach similar situations in future?“ (Z12)), hatten nur bedingten Erfolg. Hier fehlt allerdings ein größerer Datenbestand, um statistisch aussagekräftige Schlussfolgerungen treffen zu können.

Wenn man den Einfluss von nutzergenerierten Prompts auf den Text analysiert, ist festzustellen, dass es auf der einen Seite eine positive Korrelation mit dem Code für erfahrungsbasiertes Lernen gibt (siehe Tabelle 4-20). Auf der anderen Seite gibt es eine negative Korrelation mit Fragen nach Interpretation. Dies impliziert, dass nutzergenerierte Prompts zwar helfen können mehr erfahrungsbasierte Lösungen in einen Text einzubringen, aber gleichzeitig den Anteil an Fragen behindern. Es ist aber festzustellen, dass die Basis für die Berechnung mit n=21 Beiträgen vergleichsweise klein ist und daher nicht überbewertet werden darf.

**Tabelle 4-20: Korrelationen beim Einfluss von nutzergenerierten Prompts auf den Text (n=21 Beiträge, von denen in 11 Beiträgen ein Einfluss kodiert wurde;**

**13x Code SUG\_EXP und 5x Code Q\_INT)**

		<b>SUG_EXP</b>	<b>Q_INT</b>
<b>Einfluss</b>	Pearson Correlation	,603**	-,586**
	Sig. (2-tailed)	0,004	0,005

Aus Tabelle 4-16 geht hervor, dass der Prompt, der nach Lernen fragt („*What have you learned from this discussion so far?*“ (Z9, Z10)), der Prompt ist, der am häufigsten Nutzern angezeigt wurde und am häufigsten zusammen mit Beiträgen auftaucht. In Tabelle 4-21 wurde eine Korrelationsanalyse basierend auf Texten, die mit diesem Prompt zusammenhängen bzw. wo ein Einfluss kodiert wurde und dem reflexionsbezogenen Kodierschema durchgeführt. Das Ergebnis zeigt, dass der Prompt, der nach Lernen fragt, hauptsächlich mit Diskussionselementen wie Zustimmung (AGR) oder Ablehnung (DISAGR) in Verbindung steht und offenbar nicht mit Lernen selbst. Dieses Ergebnis kann vernachlässigt werden, da es in den 42 Fällen nur einen Fall gab, in dem die Kodierenden einen Einfluss der Prompts auf den Text markiert haben. Dementsprechend haben die Codes dieses einen Falles einen überproportionalen Effekt in der Analyse.

**Tabelle 4-21: Verbindung zwischen dem lernbezogenen Prompt und den reflexionsbezogenen Codes in Kroatien (n=42, von denen in einem Beitrag ein Einfluss kodiert wurde; 2x Code DISAGR und 4x Code AGR)**

		DISAGR	AGR
<b>Einfluss</b>	Pearson Correlation	,698**	,481**
	Sig. (2-tailed)	0,000	0,001

Abschließend ist festzuhalten, dass es in 40 % aller Fälle einen Einfluss des Prompts auf den Beitrag zu geben scheint (Teile des Prompts werden abgedeckt), und nur in 22 % aller Fälle wurde der Prompt komplett beachtet. Die Datenbasis für die Analyse ist allerdings nicht sehr groß, wodurch diese Ergebnisse nur vorsichtig interpretiert werden können. Die Tabelle 4-22 ist ein interessanter Ansatzpunkt, um erste Prompts (bspw. „*What have you learned from this discussion so far?*“ (Z9, Z10)) auszuschließen, da dieser sehr oft in Zusammenhang eines Textes aufgetreten ist, aber selten Einfluss auf diesen hatte.

**Tabelle 4-22: Übersicht wie oft bei einem Prompt-Einfluss auf den Text kodiert wurde**

<b>Ziele</b>	<b>Prompt</b>	<b>Keinen Einfluss</b>	<b>Teilweise Einfluss</b>	<b>Einfluss</b>	<b>Ort</b>
Z1	Try to make clear what you want to know from your colleagues	0	0	5	Neue Disk.
Z1	What did you already try to approach the situation?	3	0	1	Neue Disk.
Z4	Try to describe your point of view using a metaphor.	3	0	0	Neue Disk.
Z4	What do you think about what has happened? Do you already know how you want to approach this differently in future?	2	0	0	Neue Disk.
Z8	What would you personally suggest from your experience? Why?	0	3	3	Innerhalb Disk.
Z9, Z10	What have you learned from this discussion so far?	31	0	1	Innerhalb Disk.
Z6	Do you agree / disagree with any of the existing comments? What would you suggest from your experience? Why?	2	5	1	Innerhalb Disk.
Z12	What specific steps do you want to take to approach similar situations in future?	1	0	0	Innerhalb Disk.
Z11	Do you already have an idea how you want to tackle similar situations in future? If not, how about phrasing a new question?	2	1	1	Innerhalb Disk.
Z8	What would you personally from your experience recommend as a solution to the case and why?	3	4	2	QS-Plugin #2
Z8, Z6	What would you personally from your experience recommend as a solution to the case and why? Have you ever had a similar experience? What happened?	1	0	1	QS-Plugin #2 + #3
/	Can you please suggest the items from the house rules that you use, since we do not have the current impression of what we might be able to state, except for the obligatory preservation of office equipment of course. <sup>23</sup>	1	1	3	Nutzer
Summe:		49	14	18	



#### 4.5.5 Diskussion des Anwendungsfalles in Kroatien

In der Community of Practice in Kroatien waren ca. 30 % der Mitglieder aktiv und haben ein Thema oder eine Antwort in den Diskussionen beigetragen (siehe Tabelle 4-13). Dies ist im Vergleich zur Literatur ein üblicher Wert (Wenger, McDermott und Snyder 2002, 56). Die Community ist mit 155 Mitgliedern nicht sehr groß. Die Moderatoren sind im Vergleich zu den Nutzern überdurchschnittlich aktiv und schreiben viele neue Themen und Antworten.

Tabelle 4-14 zeigt, dass im ursprünglichen Studienzeitraum die Nutzer, wenn sie Prompts sehen, eher mehr Beiträge schreiben (pro Seitenaufruf) und eher weniger Antworten schreiben könnten (pro Seitenaufruf). Wenn der gesamte Datenerfassungszeitraum betrachtet wird, dreht sich dies um und Prompts scheinen Themen zu behindern und Antworten zu begünstigen (siehe Tabelle 4-15). An dieser Stelle ist eine sinnvolle Interpretation schwierig. Im gesamten Datenerfassungszeitraum ist der Zeitraum, in dem Prompts gezeigt wurden, anteilig deutlich kürzer. In diesem Fall scheinen die Prompts eher einen positiven Effekt auf die Häufigkeit der Erstellung neuer Antworten pro Seitenaufruf zu haben. In beiden Tabellen sind die Unterschiede zwischen der Nicht-Prompt-Kondition und der Prompt-Kondition nicht sehr groß. Dieses gegenläufige Ergebnis impliziert, dass die Prompts keinen oder nur einen sehr geringen Einfluss auf die Häufigkeit neuer Themen bzw. Antworten haben. Niedrige Effektgrößen von Prompts tauchen in anderen Studien auf (Renner u. a. 2016). Aufgrund der niedrigen Gesamtzahl an Prompts bzw. Beiträgen, die mit Prompts in Zusammenhang stehen, wurde keine Analyse des within-subjects-Designs durchgeführt, die diese Datenmenge noch stärker aufteilen würde.

Nutzergenerierte Prompts tauchten in dieser Studie ähnlich oft bei Antworten auf wie reguläre Prompts (Tabelle 4-17). Die inhaltliche Analyse ergibt jedoch, dass nutzergenerierte Antworten häufig zusammen mit den Codes für erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) auftauchen. Dies ein gewünschter Effekt für kollaborative Reflexion (siehe Abschnitt 2.5). Nachteilig ist, dass bei diesem Prompt in den Texten weniger Fragen nach Interpretationen auftauchen, die einen Beitrag zur Reflexion leisten können (siehe Abschnitt 3.4.1). Aufgrund der positiven Aspekte könnten nutzergenerierte Prompts ein interessanter Gegenstand weiterer Studien sein.

Der am häufigsten angezeigte Prompt fragt nach Lernen („*What have you learned from this discussion so far?*“; Z9, Z10; siehe Tabelle 4-16). Da dieser in der Heuristik für Situationen und Rollen in der Diskussion (Autor, Antwortender, oder Leser; siehe Tabelle

4-1) häufig war, tauchte dieser sehr oft auf. Dementsprechend gibt es viele Antworten, die geschrieben wurden, als dieser Prompt geschriebene Antworten. Eine Analyse zeigt allerdings, dass dieser in der Praxis nicht zu lernen, sondern eher zu mehr Diskussion mit Zustimmung (AGR) und Ablehnung (DISAGR) geführt hat (siehe Tabelle 4-21).

Zwei Prompts fragen nach Unerwartetem in der jüngeren Vergangenheit („*Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?*“ (Z1) und „*What was unexpected for you in your daily work in the last 2 weeks? Why?*“ (Z1)). Unerwartetes ist ein häufiger Auslöser für Reflexion (siehe Abschnitt 2.4). Beide Prompts tauchten nicht zusammen mit Antworten auf und implizieren, dass sich diese Prompts nicht dazu eignen schriftliche kollaborative Reflexion anzuregen. Die Prompts können rein gedankliche Reflexion angeregt haben, was aber aus den Daten nicht hervorgeht. Solange dies nicht messbar ist, sollte dieses Ziel zunächst als nicht unterstützbar behandelt werden.

Aus der Tabelle 4-18 geht hervor, dass 40 % der Prompts gemäß der inhaltlichen Kodierung einen Effekt auf den Text hatten. Dies kann rein zufällig geschehen sein, weil der Nutzer ohnehin in diese Richtung antworten wollte. Dennoch ist der Wert vielversprechend für weitere Studien.

Es lässt sich feststellen, dass die Prompts, die auf der Startseite angezeigt wurden, nicht funktioniert haben, da sie nie in Zusammenhang mit einem Beitrag standen (siehe Tabelle 4-17). Dies bedeutet, dass dieses Element wirkungslos war und neugestaltet werden sollte.

#### **4.6 Allgemeine Diskussion der Anwendungsfälle**

In beiden Communities wurden, sofern ein Beitrag mindestens eine Antwort bekommen hat, im Vergleich zu anderen Communities of Practice, ähnlich viele Antworten geschrieben (Jones und Churchill 2009; Kimmerle, Bientzle und Cress 2014). Dies liegt unter anderem an den Kulturen in den Organisationen. Eine ausführliche Analyse der Umgebung und der begünstigenden und behindernden Einflüsse beim Start der Communities wurde bereits publiziert (Blunk und Prilla 2017a).

Die beiden Studien unterscheiden sich dahingehend, dass nach Ablauf des ursprünglichen Studienzeitraums in Slowenien konsequent weiterhin Prompts angezeigt wurden, wohingegen in Kroatien keine Prompts angezeigt wurden. Da sich die im Folgenden beschriebenen Ergebnisse zwischen den Studien ähneln, scheint dies keinen großen Einfluss gehabt zu haben. Dies kann implizieren, dass es im Anwendungsfall in

Slowenien nicht zu Over-Prompting gekommen sein könnte (siehe Abschnitt 2.10). Die Datenlage gibt eine solche Analyse aber nicht her, da allgemein die Aktivität in der Community of Practice in Slowenien über Zeit nachgelassen hat und daher nicht genug Datenpunkte für eine umfassende Analyse zur Verfügung stehen. Eine andere Interpretation ist, dass der Effekt der Prompts in dieser Analyse überschätzt wird und es in beiden Fällen nur sehr geringe Effekte gibt und daher kein messbarer Unterschied vorhanden ist.

Bei der Analyse, wie viele Themen und Antworten in Relation zu den Lese-Events entstanden sind, ergeben die beiden Studien zusammen ein sehr unklares Bild (siehe Tabelle 4-4, Tabelle 4-5, Tabelle 4-14 und Tabelle 4-15). Daher ist eine Interpretation der Daten sehr schwierig. Letztlich hatten Prompts in den beiden Anwendungsfällen keinen eindeutigen Einfluss auf die Häufigkeit von neuen Themen oder Antworten pro Seitenaufruf. Zusätzlich sind die Unterschiede zwischen der Nicht-Prompt-Kondition und der Prompt-Kondition so gering, dass bezweifelt werden kann, dass die Prompts in diesen beiden Anwendungsfällen die Wahrscheinlichkeit erhöht haben, dass der Empfänger der Prompts einen neuen Beitrag, sei es ein neues Thema oder eine neue Antwort, verfasst hat. Diese geringen Effektgrößen wurden auch in anderen Studien beobachtet (Renner u. a. 2016).

Aus beiden Studien heraus scheint es, als seien nutzergenerierte Prompts eine vielversprechende Möglichkeit Aspekte kollaborativer Reflexion in Communities of Practice anzuregen. In der Studie in Slowenien wirkt es so, als würden nutzergenerierte Prompts mit mehr Antworten in Zusammenhang stehen (Tabelle 4-7), wohingegen in Kroatien gleiche Antwortraten zwischen regulären und nutzergenerierten Prompts beobachtet werden. Betrachtet man die inhaltliche Analyse der beiden Studien, so deuten beide Studien einen Zusammenhang zwischen nutzergenerierten Prompts und erfahrungsbasierten Lösungen (SUG\_EXP) in den verfassten Antworten an. Da dies ein wichtiges Element in kollaborativer Reflexion ist (siehe Abschnitt 2.5), macht dies nutzergenerierte Prompts für weiterführende Studien interessant. Dennoch sei erwähnt, dass die nutzergenerierten Prompts oft nicht in Zusammenhang mit neuen Antworten auftauchen (Tabelle 4-6 und Tabelle 4-16), wobei diese Prompts Nutzern nicht oft gezeigt wurden. Eine Interpretation für die Wirkung von nutzergenerierten Prompts ist, dass diese von der Formulierung her besser an den lokalen Kontext in den Organisationen angepasst sein können. Zusätzlich ist vermutlich der durch den Nutzer vorgegebene Prompt besser an das Thema angepasst als die systemgenerierten Prompts. Hier sind weitere Studien notwendig, um die Wirkung von nutzergenerierten Prompts zu analysieren.

Laut der Inhaltsanalyse scheinen die Texte, sofern Prompts gezeigt wurden, zu 40 % (Kroatien; Tabelle 4-18) und 46 % (Tabelle 4-8) teilweise oder vollständig den Prompts zu entsprechen. Dies sind hohe Werte und suggerieren, dass sich Prompts in der Praxis zum Anregen von Reflexion eignen.

Ein Beispiel, bei dem eine bestimmte Datenmenge in beiden Anwendungsfällen vorhanden ist, ist der Prompt, der nach Lernen fragt („*What have you learned from this discussion so far?*“; Z9, Z10; siehe Tabelle 4-6 und Tabelle 4-16). In der Studie in Slowenien steht dieser Prompt in Zusammenhang mit Codes für Lernen (hier: D\_LOOP) in den Texten (siehe Tabelle 4-11). In Kroatien steht der gleiche Prompt (siehe Tabelle 4-21) im Zusammenhang mit Codes für Diskussion, sprich Zustimmung (AGR) und Ablehnung (DISAGR). Es gibt aber nur einen Fall im Datensatz, bei dem die Kodierenden dem Prompt einen Einfluss beimessen. Das zeigt, dass dieser Prompt nicht universell gleich gut funktioniert und es nicht vorhersagbar ist, wie Nutzer auf diesen Prompt reagieren. Weiterhin muss unterstrichen werden, dass in beiden Anwendungsfällen die Häufigkeiten in der Analyse sehr niedrig sind und die Ergebnisse daher nur vorsichtig gewertet werden dürfen. Alle drei Auswirkungen sind wünschenswert, da somit die Diskussion angeregt wird und es zu mehr Aktivität in der Diskussion kommt. Zusätzlich hat der Prompt im Anwendungsfall in Kroatien kaum Beachtung in den Inhalten der Texte gefunden. Somit können diese Ziele höchstens als partiell unterstützt betrachtet werden. Dieser Prompt sollte weiterentwickelt werden. Für alle anderen Ziele sind nicht genug Daten vorhanden, um konkrete Aussagen treffen zu können.

In beiden Anwendungsfällen haben die Prompts auf der Startseite nicht funktioniert (Tabelle 4-7 und Tabelle 4-17), da es keinen Beitrag in beiden Studien gibt, der mit diesen Prompts in Zusammenhang steht. Dies bedeutet, dass die Funktionalität wirkungslos war. Sollte die Idee, dass auf der Startseite einer Community bereits zu Beiträgen angeregt werden soll, beibehalten werden, so muss die Funktionalität neugestaltet und wiederholt zielgerichtet erprobt werden. Im Abschnitt 6.1.9.2 wird dieses Problem intensiver diskutiert und unterschiedliche Vorschläge zur Verbesserung werden erarbeitet.

Da die Prompts für neue Themen bei dem Eingabefeld für neue Themen dargestellt wurden, welches sich unterhalb der Liste existierender Themen befindet, war dieser Prompt nicht sehr oft sichtbar. Rückblickend ist dies eine Design-Schwäche, da in Gruppen mit sehr vielen Beiträgen der Prompt selten gesehen wurde. In dem verwendeten Layout konnte der Prompt demnach nur neue Themen inhaltlich beeinflussen, aber schlecht dazu anregen neue Themen zu schreiben. Die Annahme ist, dass Nutzer, die bis zum Ende der Seite navigieren, ohnehin die Absicht verfolgen, ein neues Thema zu

verfassen. Hier sollte der Prompt prominenter platziert werden, damit Nutzer diesen auch häufiger sehen und damit der Prompt versuchen kann, mehr neue Themen anzuregen.

Beide Studien haben zudem gezeigt, dass die Prompts, die nach Unerwartetem fragen („*Can you think about a situation in your daily work in the last 2 weeks that went better than you expected? How about telling your colleagues?*“ (Z1) und „*What was unexpected for you in your daily work in the last 2 weeks? Why?*“ (Z1)), nicht in Zusammenhang mit Beiträgen aufgetaucht sind (siehe Tabelle 4-6 und Tabelle 4-16). Obwohl Unerwartetes ein typischer Auslöser für Reflexion ist (siehe Abschnitt 2.4), schaffen es dazugehörige Prompts in diesem System, zumindest keine schriftliche kollaborative Reflexion auszulösen. Ein Grund kann sein, dass die Nutzer zu dem Zeitpunkt, als sie den Prompt gesehen haben, andere Ziele zu verfolgen bzw. bei der Nutzung der Plattform nicht reflektieren wollten oder konnten. Ein Grund dafür kann mangelnde Zeit sein, da in beiden Anwendungsfällen die tägliche Arbeit mit vielen Terminen mit Klienten stark getaktet war. Eine andere Interpretation ist, dass der Prompt zu direkt formuliert war und so ggf. Nutzer abgeschreckt hat. Somit lässt sich dieses Ziel (Z1) nicht durch Prompts unterstützen. Ob es eine gedankliche individuelle oder dann mündliche kollaborative Reflexion gab, ist nicht bekannt. Hier bedarf es weiterer Studien, wie die Berater bzw. Mitglieder in der Community of Practice in der alltäglichen Praxis zur Reflexion bzw. zum Schreiben neuer Beiträge motiviert werden können. Somit kann dieses Ziel in der Praxis nicht von Prompts unterstützt werden.

#### **4.7 Einschränkungen**

Für die beiden Anwendungsfälle in Slowenien und Kroatien gibt es unterschiedliche Einschränkungen.

Bei der Studie in Slowenien wurde bereits erwähnt, dass aufgrund technischer Probleme nicht alle mit Prompts in Zusammenhang stehenden Beiträge in dem Log aufgetaucht sind. Dadurch können die Ergebnisse nur vorsichtig und ihre Aussagekraft stark eingeschränkt interpretiert werden. Da manche der Ergebnisse der Studie in Slowenien zu den Ergebnissen der Studie in Kroatien ähnlich sind, wurde die Studie bzw. wurden die Ergebnisse beibehalten.

Bei der Operationalisierung der Studie wurden Prompts als gelesen gewertet, sobald diese im Sichtfeld des Nutzers waren (siehe Abschnitt 4.3). Es ist in der alltäglichen Praxis nicht praktikabel den Nutzer bei jedem Seitenaufruf danach zu befragen, ob er den Prompt gesehen und verstanden hat. Eine andere Möglichkeit wäre die Verwendung von Eyetrackern, die Aufschluss darüber geben, ob jemand den Prompt gelesen hat. Dies ist

genauer, beantwortet aber ebenfalls nicht, ob derjenige den Prompt auch verstanden hat. Der Einsatz von Eyetrackern war allerdings in der Praxis nicht durchführbar, da mehrere Hundert Personen die Plattform genutzt haben und diese Anzahl von Eyetrackern im Projekt nicht finanzierbar gewesen wäre. Daher stellt das gewählte Vorgehen eine Näherung dar, durch die eine Studie mit der Menge an Teilnehmern möglich ist.

Bei der inhaltlichen Kodierung der Prompt-bezogenen Codes haben die Kodierenden eine Aussage darüber getroffen, inwieweit der geschriebene Beitrag, sei es Thema oder Antwort, zu dem Text des Prompts passt. Diese Übereinstimmung kann ebenfalls rein zufällig sein, falls der Nutzer ohnehin geplant hatte, bestimmte Aspekte in einem Text anzusprechen. Dieses Vorgehen stellt die bestmögliche Näherung dar.

Durch den Charakter der Feldstudie ist die Auswertung der Konditionen, in denen Nutzer Prompts gesehen haben bzw. nicht gesehen haben, sehr schwierig, da es sehr viele äußere Einflüsse gab, welche die Aktivität in der Community of Practice beeinflusst haben können. Zum einen haben die Moderatoren teilweise mehrere neue Beiträge bereitgestellt, was Nutzeraktivität zur Folge hatte. In Kroatien wurden bspw. gruppenweise neue Mitglieder eingeladen, was einzelne Aktivitätsschübe zur Folge hatte. Die Auswirkungen dieser Interventionen wurden versucht in der Auswertung so gut, wie möglich zu minimieren, was jedoch nur begrenzt möglich ist. Ebenfalls kann nicht kontrolliert werden, dass jegliche Kommunikation über die Community of Practice Plattform durchgeführt wird. Dementsprechend kann es Situationen gegeben haben, in denen jemand ein Thema anfang und nach Antworten mit anderen Personen das Thema bspw. in der Kaffeeküche weiter diskutiert hat (Prilla, Degeling und Herrmann 2012). Dementsprechend können nicht alle Aspekte der Diskussion, in denen jemand etwas gelernt hat, in der Plattform abgebildet sein.

Die Teilnehmer der Interviews haben berichtet, dass sie die Prompts bzw. die Unterstützungsfunktionen für Reflexion allgemein nicht wahrgenommen haben. Da die Prompts als gelesen gewertet wurden, sobald sie im Sichtfeld des Nutzers waren, ist es möglich, dass die Analyse die Wirkung der Prompts überschätzt und die wahrgenommenen Effekte auf anderen nicht-identifizierten Faktoren beruhen.

Ursprünglich wurde ein *within-subjects* Design geplant, was in der Analyse aufgrund des sich während der Nutzungszeit stark verändernden Umfelds anders ausgewertet wurde. Aus dem Grund gibt es in der Analyse keine verlässlichen Aussagen darüber, inwieweit die Prompts zu einem Lerneffekt geführt haben könnten.

Zusätzlich konnten viele einzelne Prompts nicht ausgewertet werden, weil die Datenmenge zu gering war. Hier wurden zu Beginn der Studie absichtlich mehr Prompts

eingefügt, um eine mögliche Ermüdung der Nutzer hinsichtlich mancher Prompts zu vermeiden. Der Nachteil ist dementsprechend, dass es weniger Daten pro Prompt gibt.

#### **4.8 Zusammenfassung**

Das in Kapitel 3 entwickelte Konzept zur Unterstützung von kollaborativer Reflexion durch Prompts wurde in zwei Anwendungsfällen in je einer Community of Practice implementiert. Die beiden Communities weisen unterschiedliche Charakteristika auf (Blunk und Prilla 2017a).

Im Rahmen der Studien Nutzer haben nur für einen bestimmten Zeitraum Prompts gesehen und vorher/anschließend nicht gesehen.

Ergebnisse der Studien zeigen, dass Prompts, die nach Unerwartetem fragen, keine Antworten bekommen haben. Das Erleben von Unerwartetem wird oft als Kernauslöser von Reflexion gesehen. Die Prompts auf der Startseite der Community of Practice Plattform wurden nicht beachtet.

Prompts schienen keinen Effekt auf die Häufigkeit neuer Beiträge gehabt zu haben.

Eine inhaltliche Analyse ergab, dass die Texte sich zu ca. 40 - 46 % teilweise oder vollständig auf die Prompts beziehen, was impliziert, dass die Nutzer die Prompts beim Verfassen neuer Beiträge beachten würden. Eine Übereinstimmung kann rein zufällig sein, da der Nutzer bestimmte Aspekte, die der Prompt abfragt, ohnehin schreiben wollte. Dennoch impliziert dieses Ergebnis, dass Prompts in der Praxis einen Effekt auf die Texte haben können.

Allgemein ist die Datenmenge in beiden Anwendungsfällen aber nicht ausreichend, um definitive Aussagen über einzelne Prompts zu tätigen.

Einer der positiven Ergebnisse ist, dass nutzergenerierte Prompts, dazu führen können, dass in den Antworten mehr erfahrungsbasierte Lösungen (SUG\_EXP) auftauchen könnten. Nutzergenerierte Prompts sind Prompts, die zuvor über eine Funktionalität im System von Nutzern ausgewählt oder geschrieben werden und dann farblich bei den übrigen Nutzern, die zum Zeitpunkt des Lesens des Themas Prompts bekommen haben, hervorgehoben werden. Zudem scheinen diese Prompts im Anwendungsfall in Slowenien mehr Antworten bekommen zu haben. Nutzergenerierte Prompts sollten in weiteren Studien untersucht werden.





## 5 Studie 2: Anregung von Reflexionselementen durch Prompts

Die Zielsetzung in dieser Studie ist die Überprüfung, inwieweit sich Prompts dazu eignen, einzelne Elemente von Reflexion anzuregen. Die Studie wurde in einer Crowdsourcing-Umgebung durchgeführt.

Diese Studie gliedert sich in drei Teilstudien: Im ersten Teil werden Prompts für Erfahrungen und Lösungen in einer Diskussion ohne Antworten verglichen. In Teil zwei werden unterschiedliche Prompts für Lösungen, die sich inhaltlich ähnlich sind, in einer Diskussion mit existierenden Antworten verglichen. Der dritte Teil dieser Studie bewertet die durch Prompts in Teilstudie 2.1 und 2.2 entstandenen Lösungsvorschläge hinsichtlich unterschiedlicher Kriterien. Die drei Teilstudien wurden zeitlich separat voneinander durchgeführt und werden dementsprechend hier getrennt berichtet.

### 5.1 Teilstudie 2.1: Prompts für Erfahrungen und Lösungen im Vergleich

Ein Ansatz, um Reflexion bestmöglich zu unterstützen, ist bestimmte Elemente bzw. Prozessschritte, wie das Austauschen von Erfahrungen oder das Entwickeln von Lösungsideen, direkt zu fördern. In dieser Teilstudie wird versucht durch Prompts direkt den Austausch von Erfahrungen und von Lösungen auszulösen und dadurch Reflexion zu begünstigen. Eine oberflächliche Beschreibung der Durchführung findet sich auch in (EmployID 2016, 31–35).

#### 5.1.1 Ansatz

Im Fokus dieser Teilstudie stehen zwei Kernelemente von kollaborativer Reflexion: der Austausch von Erfahrungen (Z6) und das Einbringen von Lösungsvorschlägen (Z8) (siehe Abschnitte 2.4 und 2.5). Diese beiden Elemente sind für Reflexion sehr wichtig. Falls es möglich ist, diese Elemente über Prompts direkt anzuregen, kann dies eine Möglichkeit bieten bspw. in der täglichen Praxis Personen dichter an Reflexion heranzuführen oder später Prompts in Systemen zu nutzen, um Personen daran zu erinnern bestimmte Aspekte in ihrer Reflexion zu beachten.

Das Ziel der Teilstudie ist zu analysieren, *inwieweit sich Erfahrungen und Lösungsvorschläge durch Prompts in einer simulierten Community of Practice Umgebung anregen lassen*. Dabei wird geprüft, ob einerseits Prompts einen Effekt auf die Antworten der Nutzer haben und andererseits, ob sich die in den Prompts angefragten Konzepte in den Antworten wiederfinden.

In einer Crowdsourcing-Umgebung ist es möglich, mit einer hohen Anzahl von Probanden und einer praxisähnlichen Situation eine Studie durchzuführen (siehe Abschnitt 3.5). Durch die hohe Anzahl von Probanden lassen sich die Daten statistisch aussagekräftig auswerten. Bei Crowdsourcing führen Personen Aufgaben mit einem sehr kleinen Umfang durch und werden dafür bezahlt. Dies erspart das Rekrutieren von Probanden für Experimente in einem Labor. Gleichzeitig können Kosten gespart werden, weil die Personen im Crowdsourcing für die geleistete Arbeit direkt bezahlt werden und so bei kleinen Aufgaben nur geringe Kosten entstehen. Eine allgemeine Darstellung von Crowdsourcing ist in Abschnitt 3.5 und eine Abbildung der vorliegenden Teilstudie innerhalb einer Crowdsourcing-Umgebung befindet sich in Abschnitt 5.1.2.

Es soll eine Diskussionsumgebung in einer Crowdsourcing-Umgebung simuliert werden, in der Probanden eine Antwort auf einen fiktiven Diskussionsbeitrag schreiben sollen und einen der Prompts angezeigt bekommen. Die Idee ist, eine Umgebung zu simulieren, die der Umgebung aus Studie 1 nahekommt (siehe Kapitel 4). Die Prompts beinhalten eine Aufforderung, eigene relevante Erfahrungen zu teilen oder einen Lösungsvorschlag einzubringen.

### 5.1.2 Methode

In diesem Abschnitt wird der Aufbau der Teilstudie beschrieben und der Einsatz in einer Crowdsourcing-Umgebung dargestellt.

#### 5.1.2.1 Umsetzung der Teilstudie

In dieser Teilstudie werden drei Prompts verglichen. Als Kontrollbedingung (KP) gibt es einen Prompt, der besagt, dass die Probanden einen Kommentar eingeben sollen („*Enter a comment*“). Dieser Prompt ist als Komponente in den beiden anderen Prompts enthalten, sodass der Unterschied zwischen dem Kontroll-Prompt und den beiden anderen Prompts dann nur der jeweilige zusätzliche Text der beiden anderen Prompts ist. Konkret sieht dies bei dem Prompt für Erfahrungen wie folgt aus: „*Have you ever been in a similar situation? **Enter a comment** that describes what you did.*“. Dieser Prompt fragt nach Erfahrungen und wird im weiteren Verlauf als „Erfahrungs-Prompt“ bzw. „EP“ bezeichnet. Damit soll der Austausch von Erfahrungen angeregt werden. Der letzte Prompt zielt darauf ab, dass der Empfänger Lösungsvorschläge (LP) einbringen soll und lautet „*Enter a comment. Describe a solution to the problem above and why you think this is a solution*“. Weitere Unterschiede zwischen Kontroll-Prompt und Lösungs- bzw. Erfahrungs-Prompt gibt es nicht. In Tabelle 5-1 ist eine Übersicht der Prompts.

Tabelle 5-1: Prompts in Teilstudie 2.1

Name	Abkürzung	Prompt
Erfahrungs-Prompt	EP	Have you ever been in a similar situation? <b>Enter a comment</b> that describes what you did
Lösungs-Prompt	LP	<b>Enter a comment.</b> Describe a solution to the problem above and why you think this is a solution
Kontroll-Prompt	KP	<b>Enter a comment</b>

Für die Teilstudie wird ein *between-subjects-Design* ausgewählt, was bedeutet, dass jeder Proband nur eine Experiment-Kondition sieht. Dies dient zur Vermeidung von Redundanzen und Übertragungseffekten. Da sich jeder Proband nur in einer Experiment-Kondition befindet, folgt daraus, dass jeder Proband nur einen der Prompts sieht.

### 5.1.2.2 Abbildung in Crowdfunder

Die Teilstudie wurde in der Plattform Crowdfunder (heute FigureEight) durchgeführt. Bei der Erstellung einer Studie in einer Crowdsourcing-Umgebung gibt es unterschiedliche Aspekte, wie bspw. Qualitätssicherung, zu beachten. Zusätzlich wird in diesem Abschnitt die simulierte Forenumgebung dargestellt.

#### 5.1.2.2.1 Simulierte Forenumgebung

Das Ziel war es, eine Diskussionssituation innerhalb eines Jobs zu simulieren, sodass die Probanden das Gefühl haben, sie würden sich untereinander zu Herausforderungen und Problemen in ihrer Tätigkeit austauschen können. Diese Diskussionssituation sollte so ähnlich aussehen, wie ein herkömmliches Forum, damit es der Situation in der ersten Studie (siehe Kapitel 4) ähnelt. Das einem Forum nachempfundene Layout ist in Abbildung 5-1 dargestellt.

Die Probanden sollten für den Beitrag (Abbildung 5-1, (2)), der ein typisches Problem für einen Crowdfunder beschreibt, in einer Textbox weiter unten eine Antwort verfassen. Bei der Erstellung von Crowdsourcing-Jobs ist es wichtig, dass es für die Crowdfunder konkrete Instruktionen gibt (1). Hier steht zwar, dass die Probanden sich an die Instruktionen halten und dem Hinweis folgen sollen. Dies bedeutet nicht, dass die Probanden das automatisch tun. Es gab allerdings keine Strafe, falls sie Instruktionen nicht beachtet haben. Einzig Antworten, bei denen sich der Ersteller keine Mühe gegeben hat, wie bei aus Wikipedia kopierte Texte, wurden nicht akzeptiert.

In der Abbildung 5-1 ist ebenfalls das zu beantwortende Problem (2) gekennzeichnet. Das abgebildete Problem beschreibt einen Crowdfunder, der negatives Feedback auf seine Jobs erhalten hat, was auf der Plattform eine ausbleibende Bezahlung nach sich zieht.

Daraus resultierte für den fiktiven Crowdworker ein Motivationsproblem. Dies ist ein typisches Problem von Crowdworkern, was aus der Literatur entnommen wurde (bspw. Martin u. a. 2014). Da Crowdworker sich oft in Communities organisieren und sich austauschen, ist dies ein passendes Szenario. In dieser Teilstudie haben alle Probanden das gleiche Szenario bekommen.

*Read the following text carefully and be sure to understand what it is about. You are supposed to create a comment on the text later in this job.* **1**

Sometimes I get negative feedback on jobs I completed. Then the requester rejects my work and refuses payment although I did my best to answer the questions correctly and truthfully. This happened to me twice last week. Afterwards I get problems motivating myself because I invested so much time and effort. So I sometimes ask myself why should I take more jobs if I always get such negative experiences.

*How many times did the problem happen to the author last week?* **2**

☐ Once (1)  
☐ Twice (2)  
☐ Three times (3)  
☐ More than five (5) times

*Please provide your comment in the text box below. Be sure to write at least three full sentences and to adhere to the statement written above the comment field. Good comments will receive a bonus.* **3**

**Enter a comment** **4**

**1**

❗ Create the comments from what comes into your mind. Do not make up anything but just write what you associate with the text you read and the statements above the text box.

**Abbildung 5-1: Simulierte Forenumgebung in Crowdfower mit einem Thema, zu dem eine Antwort verfasst werden soll. Es wird im Bild des Kontroll-Prompts angezeigt.**

#### 5.1.2.2.2 Qualitätssicherung

Aus den Abschnitten 3.6.7 und 3.6.9 ergeben sich unterschiedliche Anforderungen an das Design von Crowdsourcing-Studien. In den folgenden Abschnitten wird beschrieben, wie die Teilstudie in der Umgebung in Crowdfower aufgesetzt wurde.

Um das between-subjects-Design in Crowdfower abzubilden, wurde der Job so konfiguriert, dass jede Person diesen nur ein einziges Mal durchführen durfte. Dies hat

gleichzeitig den positiven Nebeneffekt, dass Nutzer mit Täuschungsabsicht nur in geringem Umfang unpassende Antworten geben können. In die Teilstudie wurden zufällig die ersten 300 Personen aufgenommen, die den Job durchgeführt haben. Dementsprechend gibt es 100 Ergebnisse pro Prompt-Kondition. Um mögliche Zeitversatz-Effekte zu vermeiden, wurde der Prompt pseudo-zufällig ausgewählt, sobald ein Crowdworker den Task geöffnet hat. Zusätzlich war der Job nur zu bestimmten Tageszeiten aktiv (9:00 – 21:00 CET).

Crowdflower hat intern drei Ränge an Qualität von Crowdworkern, aus denen ein Ersteller eines Jobs wählen kann. Je höher der Rang ist, desto höher sei die entsprechende geleistete Qualität der Arbeit, aber desto länger kann es dauern, bis ein Job abgeschlossen ist, weil weniger Personen mit diesem Rang verfügbar sind. Gleichzeitig ist die Wahrscheinlichkeit niedriger, dass Crowdworker mit Täuschungsabsicht unter den Probanden sind.

Zur Gewährleistung der Qualität wurden verschiedene Maßnahmen implementiert. Zum einen mussten die Personen eine Testfrage beantworten, die sich inhaltlich aus der simulierten Diskussion ergeben hat (siehe Abbildung 5-1 (3)). Dies ist eine Gold-Standard-Frage (siehe Abschnitt 3.6.7), weil die Antwort auf die Frage vorher bekannt ist. Dies hilft sicherzustellen, dass der Crowdworker den Text zumindest gelesen hat. Personen, die diese Frage falsch beantworten, können somit leicht aus der Teilstudie ausgeschlossen werden. Die Frage zielt darauf ab, wie oft der Autor des fiktiven Problems bereits in dieser Situation war (Antwort in dieser Studie: zwei Mal).

Weiterhin wurde zur Vermeidung von Sprachbarrieren der Job so konfiguriert, dass nur Personen aus bestimmten Ländern darauf zugreifen konnten. Da der Job auf Englisch ausgeschrieben wurde, wurden nur nativ englischsprachige Länder zugelassen (hier Großbritannien und USA), sowie Länder, die zum Zeitpunkt der Teilstudie die höchsten Bewertungen im *English Language Proficiency Index*<sup>24</sup> hatten. Maximal konnten bei Crowdflower zwölf Länder ausgewählt werden. Diese sind: USA, Großbritannien, Deutschland, Polen, Kroatien, Norwegen, Niederlande, Schweden, Slowenien, Belgien, Österreich, Finnland, Estland und Dänemark.

Für die Bezahlung wurde darauf geachtet, dass die Crowdworker fair bezahlt werden. Es wurden 0,80 \$ für die Teilnahme gezahlt, was auf einem Stundenlohn von acht US-Dollar pro Stunde basiert (Durchführung der Teilstudie November 2015). Somit wird erwartet,

---

<sup>24</sup> <https://www.ef.com/wwen/epi/> (Letzter Zugriff am 12.08.2019)

dass die Beantwortung des Tasks ca. sechs Minuten in Anspruch nimmt. Zusätzlich wurde für sehr gute Antworten ein Bonus von 0,20 \$ ausgezahlt.

#### **5.1.2.3 Probelauf**

Bevor der Versuch durchgeführt wurde, gab es einen Probedurchlauf mit 20 Crowdworkern, um zu prüfen, ob der Job nachvollziehbar ist. Die Probanden haben angegeben, dass sie sich mit dem fiktiven Problem identifizieren konnten und manche der Beiträge der Teilnehmer waren geschrieben, als wäre die Forenumgebung real: „*If you believe that you were treated unfairly, you should report it to CrowdFlower. (...) We should all strive for improving it together. It is important not to let negative experiences discourage us from trying again.*”. Hier wird explizit der fiktive Autor des Themas in der Forenumgebung angesprochen. Es haben zwar nicht alle Teilnehmer so geantwortet, aber dieses Beispiel ist repräsentativ für den Großteil der Kommentare.

Zusätzlich wurde bei diesem Test die Bezahlung als fair bewertet.

#### **5.1.2.4 Methodik für die Analyse**

Um zu evaluieren, ob die Prompts Einfluss auf den Inhalt der Antworten genommen hat, bedarf es einer qualitativen Inhaltsanalyse. Dazu wird das bereits beschriebene reflexionsbezogene Kodierschema (siehe Abschnitt 3.4) genutzt.

Jede Antwort wurde von zwei Personen kodiert. Da beide Kodierende bereits Erfahrung mit dem Kodierschema haben, gibt es für diese Teilstudie keine intensive Trainingsphase, sondern einen kurzen Testlauf, welche die Kodierung von 20 Antworten umfasst. Nach einer Diskussion über die Anwendung des Schemas kodiert jeder Forscher den kompletten Datensatz. Der Autor dieser Arbeit hat den Datensatz mitkodiert.

Um zu überprüfen, ob andere Faktoren, wie Alter, Bildungsgrad, Geschlecht oder Erfahrung als Crowdworke**r** bzw. wöchentliche Arbeitszeit als Crowdworke**r** einen Einfluss haben, werden innerhalb des Jobs zusätzlich demografische Daten abgefragt. Diese Daten können die Crowdworke**r** freiwillig angeben. Zusätzlich erfasst Crowdflower automatisch das Herkunftsland. Dies ermöglicht eine Überprüfung, ob dieses einen Einfluss hat. Da man über VPN oder Verschleierung des DNS seine Herkunft im Internet verschleiern kann, ist bei diesen Daten die Aussagekraft vergleichsweise gering.

### 5.1.3 Hypothesen

Die Grundannahme der Teilstudie ist, dass Prompts eine Auswirkung auf die Antworten der Probanden haben können (siehe 2.10). Zur besseren Analyse der Forschungsfrage in dieser Teilstudie wurden unterschiedliche Hypothesen entwickelt. Die Forschungsfrage lautet: *Inwieweit lassen sich Erfahrungen und Lösungsvorschläge durch Prompts in einer simulierten Community of Practice Umgebung anregen?*

Wenn Prompts an sich eine Auswirkung auf die Texte der Nutzer haben, dann sollte ein Prompt für Erfahrungen zu einem höheren Anteil von Erfahrungen in den Antworten führen (Z6).

*Hypothese 1: Antworten, die durch die Erfahrungs-Prompts (EP) entstehen, werden mehr Erfahrungen beinhalten als die Antworten, die durch die Lösungs-Prompts (LP) oder Kontroll-Prompts (KP) entstehen.*

Um diese Hypothese und die weiteren Hypothesen beibehalten zu können, muss die jeweilige Null-Hypothese widerlegt werden. Dazu muss gezeigt werden, dass es einen Unterschied zwischen den Konditionen in der Inhaltsanalyse gibt (siehe Abschnitt 3.4). Da Hypothese 1 auf erfahrungsbasierte Inhalte abzielt, wird dazu der Code für Erfahrung (EXP) aus dem reflexionsbezogenen Kodierschema herangezogen. Die Hypothese wird wie folgt operationalisiert:  $EP(EXP) > LP(EXP), KP(EXP)$ .

Ähnlich wie bei dem erfahrungsbasierten Prompt versucht der lösungsorientierte Prompt die Empfänger dazu zu motivieren, dass mehr Lösungen geschrieben werden (Z8).

*Hypothese 2: Antworten, die durch die Lösungs-Prompts (LP) entstehen, werden mehr Lösungen enthalten als die Antworten, die durch die Erfahrungs-Prompts (EP) oder Kontroll-Prompts (KP) entstehen.*

Die Operationalisierung des zweiten Prompts basiert auf den Codes für Lösungsvorschläge, die entweder erfahrungsbasiert (SUG\_EXP) oder wissensbasiert (SUG\_KNO) sein können. Beide werden als allgemeine Lösungsvorschläge (SUG) zusammengefasst. Die Hypothese kann beibehalten werden, wenn:  $LP(SUG) > EP(SUG), KP(SUG)$ .

Um eine konkretere Analyse zu ermöglichen, wird die Hypothese feiner aufgegliedert. Erfahrungsbasierte Lösungsvorschläge sind für die Reflexion wertvoller als wissensbasierte Vorschläge, da erfahrungsbasierte Aussagen zusammen in der Reflexion mit Lernen auftreten (Prilla u. a. 2015). Die Vermutung ist, dass dies insbesondere auftritt, wenn die Prompts nach Erfahrungen fragen.

*Hypothese 3: Antworten, die durch die Erfahrungs-Prompts (EP) entstehen, werden mehr erfahrungsbasierte Lösungsvorschläge beinhalten als die Antworten, die durch die Lösungs-Prompts (LP) oder Kontroll-Prompts (KP) entstehen.*

Diese Hypothese lässt sich durch die existierenden Codes für erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) operationalisieren:  $EP(SUG\_EXP) > LP(SUG\_EXP)$ ,  $KP(SUG\_EXP)$ . Wissensbasierte Lösungsvorschläge werden nicht vermutet.

Unbegründete Hinweise sind Vorschläge, die keine Begründung aufweisen, wie es bspw. bei den erfahrungs- oder wissensbasierten Vorschlägen der Fall ist. Diese sind in der Reflexion nicht als hilfreich angesehen, da Reflexion erfahrungsbasiertes Lernen ist (siehe Abschnitt 2.4). Die Prompts sind so gestaltet, dass ein Lösungsvorschlag mit Begründung gegeben wird, oder eine Erfahrung mit einer Handlungsbeschreibung eingebracht werden. Beides sollte eine Begründung liefern und dementsprechend nicht zu unbegründeten Vorschlägen führen. Folglich müsste der Kontroll-Prompt am meisten unbegründete Vorschläge nach sich ziehen.

*Hypothese 4: Antworten, die durch die Kontroll-Prompts (KP) entstehen, werden mehr unbegründete Hinweise beinhalten als die Antworten, die durch die Erfahrungs-Prompts (EP) oder Lösungs-Prompts (LP) entstehen.*

Dieser Prompt lässt sich über den Code für Hinweise (ADV) operationalisieren. Die Hypothese kann behalten werden, wenn  $KP(ADV) > EP(ADV)$ ,  $LP(ADV)$ .

#### **5.1.4 Deskriptive Analyse der Daten**

Eine Beschreibung der Vorgehensweise für die statistische Analyse und die Überprüfung der Daten auf eventuelle Voraussetzungen für statistische Verfahren ist in Anhang A.

Bevor mit der Analyse der Daten begonnen werden kann, müssen die Daten bereinigt werden. Von insgesamt 300 Antworten (100 je Prompt) wurden 278 Antworten behalten. Es galten zusätzlich die folgenden Kriterien: Die Antwort muss mindestens drei Sätze enthalten und es muss ein Bezug auf das beschriebene Problem geben. Bei Nichtbeachtung dieser Regeln wurden die geschriebenen Texte aus der Auswertung ausgeschlossen.

Insgesamt 15 Personen haben die Gold-Standard-Frage falsch beantwortet. Nach manueller Durchsicht dieser Kommentare stellte sich aber heraus, dass diese Personen dennoch den Text ehrlich und umfassend beantwortet haben. Dadurch wurden diese



Antworten für die Analyse beibehalten. Dies zeigt auch, dass es nur einen eher geringen Anteil von Nutzern mit möglicher Täuschungsabsicht gab.

Im Folgenden wird geprüft, ob die demografischen Daten einen Einfluss in der Teilstudie auf die für die Hypothesen relevanten reflexionsbezogenen Codes haben: Erfahrung (EXP), allgemeine Vorschläge (SUG), erfahrungsbasierte Vorschläge (SUG\_EXP) und Hinweise (ADV). Neben der Referenz auf Erfahrung finden sich in solchen Datensätzen oft Referenzen auf Wissen (KNO) und wissensbasierte Lösungsvorschläge (SUG\_KNO).

**Tabelle 5-2: Übersicht der deskriptiven Daten in Teilstudie 2.1**

		<b>Kontroll-P.</b>	<b>Erfahrungs-P.</b>	<b>Lösungs-P.</b>	<b>Alle</b>
	Antworten nach Bereinigung	94	92	92	278
<b>Zeit</b>	Durchschnitt	0:08:23	0:08:26	0:09:03	0:08:37
	Median	0:07:00	0:07:27	0:07:31	0:07:19
	St. Abw.	0:05:39	0:04:58	0:05:47	0:05:29
	Min.	0:02:05	0:02:03	0:01:27	0:01:27
	Max	0:25:50	0:27:57	0:29:33	0:29:33
<b>Geschlecht</b>	Männlich	50	50	50	150
	Weiblich	43	42	42	127
	Keine Antwort	1	0	0	1
<b>Alter</b>	Durchschnitt	43,3	37,9	30,6	37,4
	Median	42,5	35,0	29,0	35,0
	St. Abw.	11,7	11,0	8,5	11,8
	Min.	19	19	18	18
	Max	71	67	59	71
	Keine Antwort	0	6	2	8
<b>Gold-Standard-Frage</b>	Richtige Antwort	88	87	88	263
	Falsche Antwort	6	5	4	15
<b>Wörter</b>	Anzahl Wörter	6399	7036	6944	20379
	Durchschnitt	68,07	76,48	75,48	73,31
	Median	66,50	64,50	60	64
	St. Abw.	44,24	52,84	60	44,98
	Min.	20	14	11	11,00
	Max	293	332	182	332,00

Aus der Tabelle 5-2 geht hervor, dass die Probanden durchschnittlich 8:37 Minuten für die Erledigung des Jobs gebraucht haben. Das liegt knapp über den veranschlagten sechs Minuten. Die Bezahlung des Jobs wurde von den Probanden insgesamt als gut bewertet. Es muss aber unterstrichen werden, dass diese Zeit nur den Zeitraum zwischen dem Öffnen und dem Abschicken der Antwort beinhaltet. Eventuelle Ablenkung der Nutzer, bspw. durch das Smartphone, wurde nicht erfasst. Die Zeit zwischen allen drei Konditionen ist vergleichbar. Diese Werte wurden aufgrund der Streuung in eine neue

Variable gruppiert: 0-240s, 241-480s, 481-720s, und mehr als 720s. Wie viele Werte jeder Gruppe zu sortiert wurden, ist in Tabelle 5-3 abgebildet. Ein  $\chi^2$ -Test ergab, dass die neue Variable mit Zeit-Gruppen in keinem Zusammenhang mit einem reflexionsbezogenen Code steht.

**Tabelle 5-3: Anzahl der Werte für jede Gruppe in der gruppierten Variable für die Zeit**

Gruppe	Anzahl Werte
<b>0-240s</b>	38
<b>241-480s</b>	120
<b>481-720s</b>	70
<b>&gt;720s</b>	49

Die Werte für die Anzahl der Wörter sind ebenfalls grob vergleichbar, wobei die Texte für den Kontroll-Prompt durchschnittlich sieben Wörter kürzer sind als die Texte für die anderen beiden Prompts. Die Standardabweichung der Werte ist bei dem Lösungs-Prompt am größten, wobei die Länge des längsten Beitrags deutlich kürzer ist als bei den übrigen Prompts. Die Werte für den Durchschnitt bzw. Median und die Standardabweichung zeigen, dass die Daten insbesondere am oberen Ende sehr streuen. Da die Daten stark streuen, wurden die Werte in einer neuen Variable gruppiert (Variable *WordCount*): 0-40, 41-80, 81-120 und mehr als 121 Wörter. In den Gruppen sind 88, 134, 41 bzw. 15 Werte enthalten. Ein Clusterverfahren von SPSS wurde nicht eingesetzt, da dieses in Testläufen mehrere Cluster mit 2-3 Einträgen produziert hat, was sich aufgrund der kleinen Gruppengrößen nicht für die folgenden statistischen Verfahren eignet. Ein  $\chi^2$ -Test zeigt, dass es keinen signifikanten Zusammenhang mit dem Code für Hinweise (ADV) gibt. Alle anderen betrachteten Codes zeigen im  $\chi^2$ -Test einen signifikanten Zusammenhang (siehe Tabelle 5-4).

**Tabelle 5-4: Anteil wie oft der Code einem Text je nach Anzahl der Wörter hinzugefügt wurde**

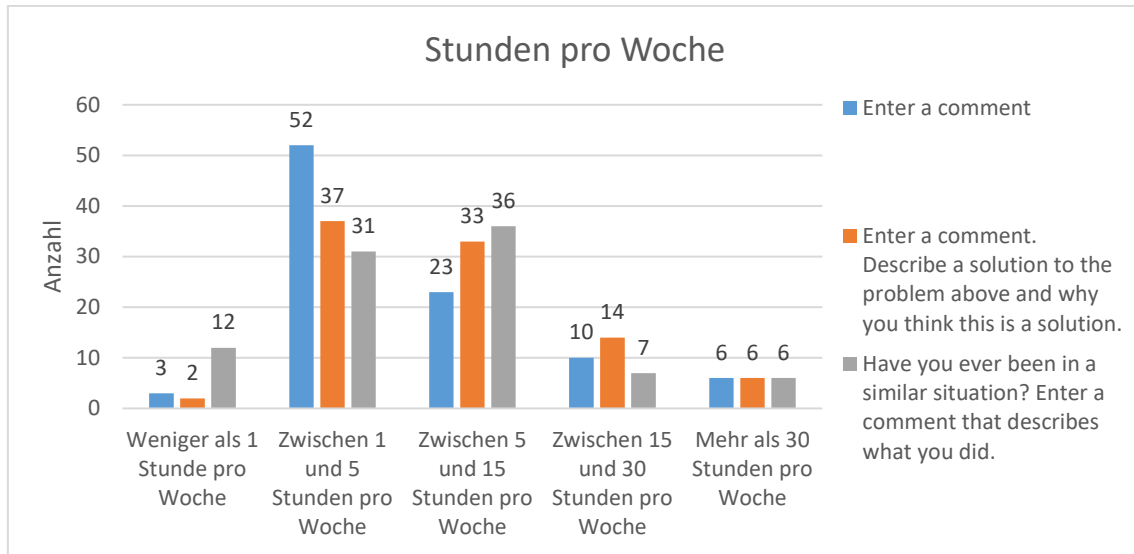
Code	0-40	41-80	81-120	121+	$\chi^2$ -Test
<b>EXP</b>	49,2 %	50,8 %	69,1 %	76,5 %	$\chi^2(3)=12,090$ , $p=,007$
<b>KNO</b>	42,6 %	68 %	72,7 %	76,5 %	$\chi^2(3)=17,151$ , $p=,001$
<b>SUG</b>	36,1 %	59,4 %	74,5 %	91,2 %	$\chi^2(3)=33,384$ , $p<,001$
<b>SUG_EXP</b>	9,8 %	25 %	36,4 %	50 %	$\chi^2(3)=20,960$ , $p<,001$
<b>SUG_KNO</b>	26,2 %	39,1 %	49,1 %	55,9 %	$\chi^2(3)=10,300$ , $p=,016$

Die Tabelle 5-4 zeigt, wie oft je nach Anzahl der Wörter, der entsprechende Code vergeben wurde. Beispielsweise wurde der Code für Erfahrung (EXP) bei Antworten mit 0-40 Wörtern zu einem Anteil von 49,2 % vergeben (und impliziert, dass 50,8 % der Beiträge mit 0-40 Wörtern keine Erfahrung beinhalten). Die Verteilung der Codes über die Gruppierungen der Wörter impliziert in dieser Tabelle, dass je mehr Wörter geschrieben werden, desto höher der Anteil der Codes ist bzw. desto eher wurde der Code

in einem Text vergeben. Über diese Tabelle lässt sich ablesen, wie viele Wörter den größten anteiligen Anstieg haben. Bei dem Code Erfahrung bspw. ist der Unterschied zwischen 0-40 Wörtern und 41-80 Wörtern vernachlässigbar. Wenn jedoch ca. 40 Wörter mehr geschrieben werden (Sprung von 41-80 zu 81-120), dann ist der Anteil der Text, in denen ein Code für Erfahrung enthalten ist, um knapp 20 % größer. Im Vergleich zwischen 81-120 und mehr als 120 Wörtern ist der Anstieg der Texte, die einen Code für Erfahrungen beinhalten nicht mehr so groß. Dies deutet darauf hin, dass eine minimale Anzahl an Wörtern notwendig ist, um eine Erfahrung zu transportieren, und es ein Limit gibt, ab wann dies erreicht ist. Bei den Werten für erfahrungsbasierte Vorschläge zeigt sich ein kontinuierlicher Anstieg über die Gruppen der Wörter hinweg. Der Code für allgemeine Vorschläge (SUG) umfasst dabei die Menge von sowohl erfahrungsbasierten (SUG\_EXP) als auch wissensbasierten Lösungsvorschlägen (SUG\_KNO).

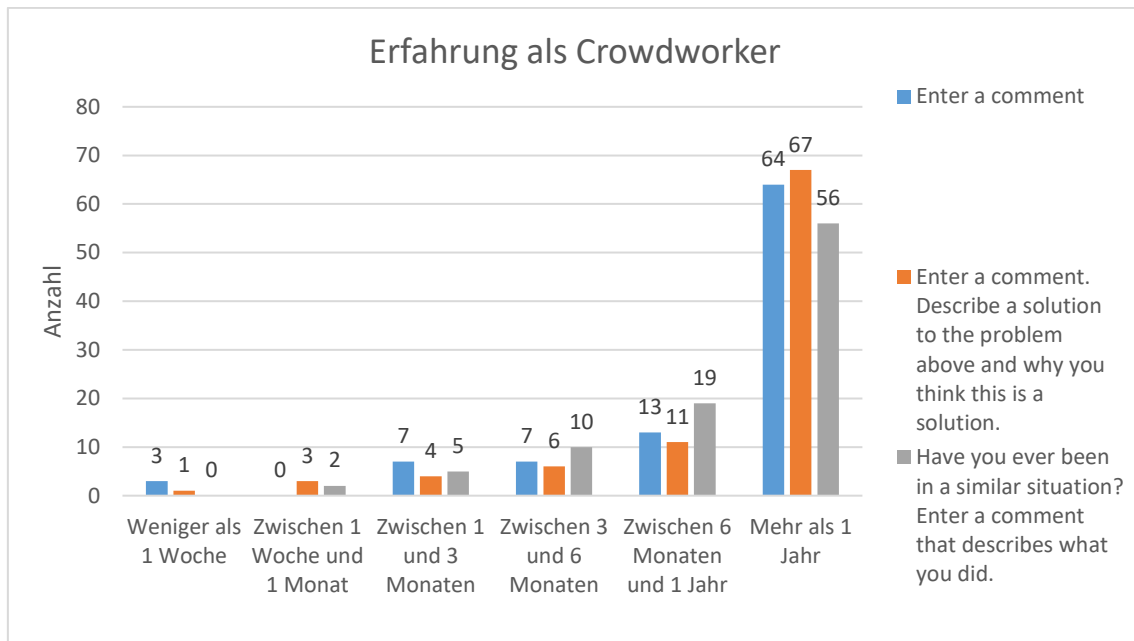
Insgesamt haben den Job knapp mehr Männer als Frauen erledigt (54 % zu 46 %), was darauf hindeutet, dass die Demografie der Probanden leicht von bspw. der Demografie auf Mechanical Turk abweicht, wo ältere Umfragen bis zu 70 % weibliche Crowdfworker ermittelt haben (Ipeirotis 2010). Jüngere Umfragen gehen von 52 % Frauenanteil aus (Ross u. a. 2010). Eine Person hat das Geschlecht nicht angegeben. Das Geschlecht ist in allen Konditionen gleich verteilt. Ein  $\chi^2$ -Test ergab, dass das Geschlecht in keinem Zusammenhang mit den reflexionsbezogenen Codes steht.

Im Vergleich zu den Studien zu Demografien von Nutzern auf Mechanical Turk sind in diesem Datensatz die Probanden im Schnitt leicht älter (Ipeirotis 2010; Ross u. a. 2010). Acht Nutzer haben ihr Alter nicht angegeben. Die Probanden waren zwischen 18 und 71 Jahre alt. Das Alter der Probanden hat keinen Einfluss auf die Antworten. Für die Kalkulation wurden die Probanden in Altersgruppen (18-24, 25-34, 35-44, 45-54, 55-64, 65+) eingeteilt, wie bspw. bei Ross et al. (2010). Ein  $\chi^2$ -Test zeigte keinen Zusammenhang zwischen den Altersgruppen und den reflexionsbezogenen Codes.



**Abbildung 5-2: Eigene Auskunft zur wöchentlichen Arbeitszeit als Crowdworker**

Abbildung 5-2 zeigt, wie viele Stunden die Crowdworker durchschnittlich pro Woche laut eigener Aussage im Crowdsourcing tätig sind. Auffällig ist, dass es wenige Personen gibt, die weniger als eine Stunde pro Woche und mehr als 30 Stunden arbeiten. Die meisten Crowdworker investieren zwischen einer und 15 Stunden pro Woche und wenige liegen bei 15 bis 30 Stunden pro Woche. Eine Interpretation ist, dass in der Population dieser Teilstudie wenig Personen sind, die Crowdsourcing als Vollzeit-Job bzw. einzige Einnahmequelle betrachten. Die wöchentliche Arbeitszeit ist unter den Konditionen ähnlich verteilt, wobei es bei den niedrigeren wöchentlichen Arbeitszeiten größere Schwankungen gibt. Laut einem  $\chi^2$ -Test gibt es keinen Zusammenhang zwischen den durchschnittlich pro Woche geleisteten Stunden und den reflexionsbezogenen Codes. Die Gruppe für „Weniger als 1 Stunde pro Woche“ wird jedoch (bis auf den Erfahrungs-Prompt) nicht berücksichtigt, da die Anzahl der Werte kleiner als fünf ist, was für einen  $\chi^2$ -Test benötigt wird.



**Abbildung 5-3: Eigene Auskunft zur Arbeitserfahrung als Crowdworker in Teilstudie 2.1**

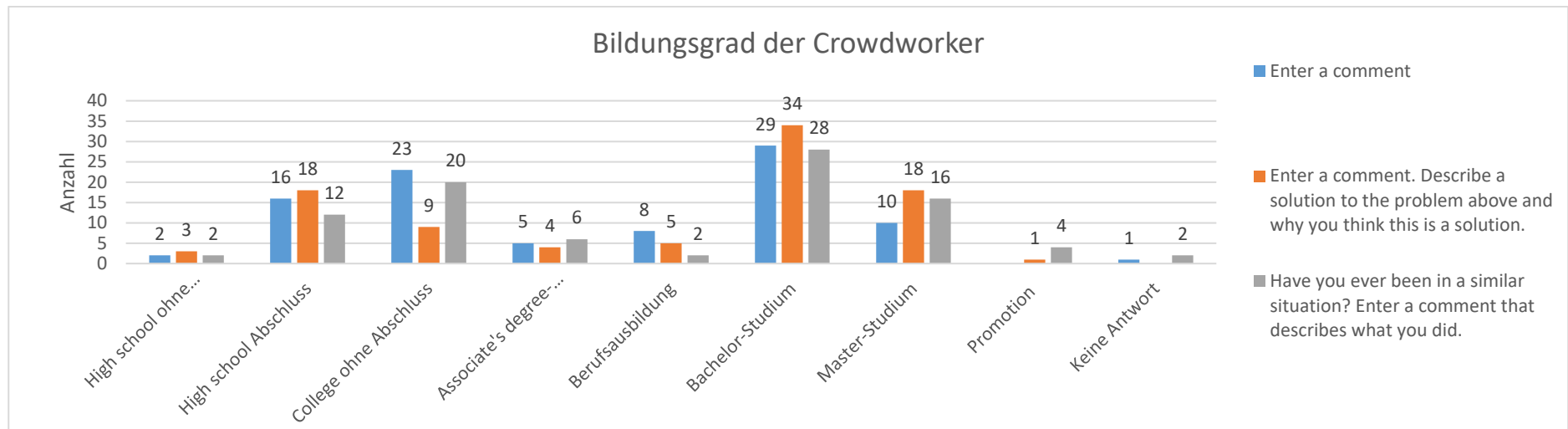
Die Crowdworker in dieser Teilstudie hatten mit großer Mehrheit über ein Jahr Erfahrung als Crowdworker, wie Abbildung 5-3 zeigt. Diese Verteilung der Erfahrung ist zu erwarten gewesen, da für das Experiment nur Crowdworker mit dem höchsten Rang ausgewählt wurden. Dies ist nicht leicht zu erreichen und dementsprechend finden sich nur sehr wenige Crowdworker in dem Datensatz, die bspw. weniger als drei Monate Erfahrung besitzen. Die Erfahrung ist über die Konditionen relativ gleich verteilt. Da viele der Crowdworker mehr als ein Jahr auf der Plattform aktiv waren und bei den anderen Kategorien die Anzahl der Werte oft zu klein sind, um in einem  $\chi^2$ -Test berücksichtigt zu werden, wurde eine Variable erstellt, welche die Crowdworker anhand dieser Anforderung in zwei Gruppen geteilt hat: mehr als ein Jahr Erfahrung und weniger als ein Jahr Erfahrung (Variable: *ExpMoreThanAYear*). Es zeigte sich in einem  $\chi^2$ -Test kein Zusammenhang zwischen der neuen Variable und den reflexionsbezogenen Codes.

Im Wesentlichen lassen sich die Crowdworker in dieser Teilstudie anhand des Bildungsgrads in zwei Gruppen einsortieren (Abbildung 5-4): Eine Gruppe hat einen High-School-Abschluss und anschließend (noch) keinen College-Abschluss erzielt. Die andere Gruppe hat studiert und einen Bachelor oder Masterabschluss erreicht. Allgemein sind in mehreren Kategorien die Anzahlen für einen  $\chi^2$ -Test zu gering. Eine Gruppierung der Crowdworker in Akademiker und Nicht-Akademiker zeigte in einem  $\chi^2$ -Test keinen Zusammenhang (Variable: *Academics*).

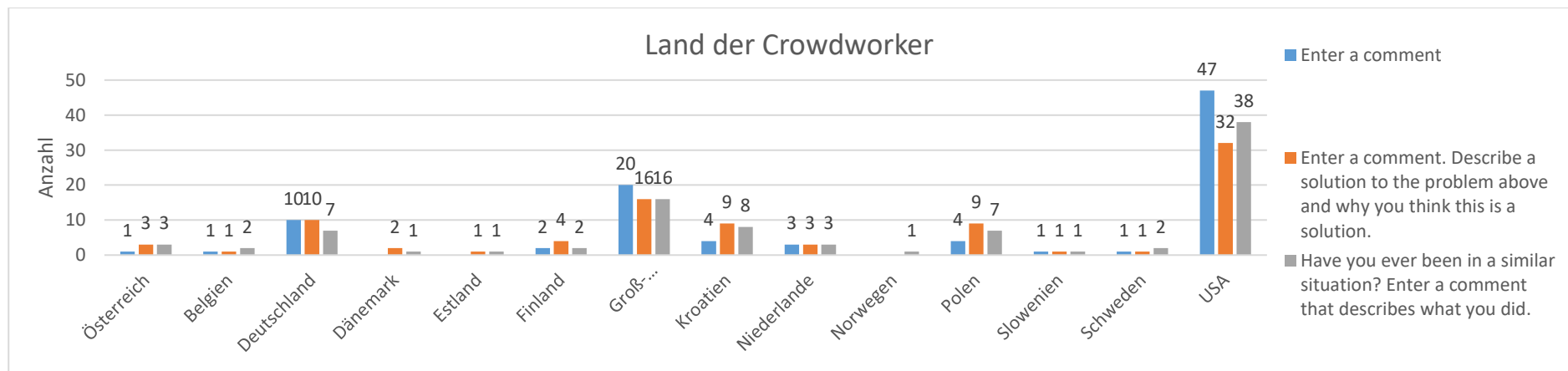
Abbildung 5-5 zeigt eine Übersicht, aus welchem Land die einzelnen Crowdworker laut Crowdfunder stammen. Es zeigt sich, dass die überwiegende Mehrheit der Crowdworker

aus den USA und aus Großbritannien stammen. Hier erfüllen viele Länder nicht die Voraussetzungen für einen  $\chi^2$ -Test, da pro reflexionsbezogenem Code ein  $n \geq 5$  je Land vorhanden sein muss. Im Folgenden werden die Länder danach gruppiert, ob deren Bürger Englisch als Muttersprache haben (Variable: *NativeSpeaker*; Gruppe 1: Großbritannien und USA; Gruppe 2: übrige Länder). Dies ist eine Verallgemeinerung, da die Daten von Crowdfower stammen und keine Selbstauskunft der Teilnehmer sind. Eine Person, die bspw. kürzlich aus einem nicht-englischsprachigen Land in die USA zugezogen ist und dort Crowdsourcing betreibt, fällt dennoch in die Kategorie *NativeSpeaker*. Diese Annäherung scheint für diese Analyse plausibel, da zumindest der überwiegende Teil der Personen länger als ein Jahr Crowdsourcing betreibt und daher angenommen werden kann, dass Personen, die nicht-muttersprachlich Englisch aufgewachsen sind, darüber Englisch lernen konnten (obwohl es Jobs gibt, bei denen Fähigkeiten in der englischen Sprache benötigt werden).

Ein  $\chi^2$ -Test zeigt, dass die Variable *NativeSpeaker* Einflüsse hat. Diejenigen, die Englisch als Muttersprache sprechen, haben 71 % der Hinweise (ADV;  $\chi^2(1)=5.609$ ,  $p=,018$ ) und 75 % der erfahrungsbasierten Lösungsvorschläge (SUG\_EXP;  $\chi^2(1)=8.296$ ,  $p=,004$ ) geschrieben. Bei Erfahrungen (EXP) oder wissensbasierten Vorschlägen (SUG\_KNO) gibt es keinen signifikanten Unterschied. An dieser Stelle ist eine Interpretation schwierig, da einerseits erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) bei Personen auftauchen, die Englisch als Muttersprache sprechen, aber andererseits keine auffallende Menge an Erfahrungen beigetragen wird (EXP). Der Vermutung, dass es den Personen leichter fällt, Lösungen zu formulieren, widerspricht ein fehlender Zusammenhang zu dem Code für wissensbasierte Lösungen (SUG\_KNO).



**Abbildung 5-4: Eigene Auskunft zum Bildungsgrad der Probanden in Teilstudie 2.1**



**Abbildung 5-5: Angabe der Länder aus denen die Crowdworker aus Teilstudie 2.1. Angaben von Crowdflower.**

### 5.1.5 Qualitative Analyse

In diesem Abschnitt werden die einzelnen Hypothesen nacheinander geprüft. Erste Kalkulationen von Krippendorffs Alpha haben eine moderate Übereinstimmung ergeben (Hayes und Krippendorff 2007). Nach üblichen Ansätzen wurden Abweichungen diskutiert und jeder Kodierende konnte die eigene Kodierung nach Bedarf anpassen (Johri und Yang 2017; Hamann 2002). Das durchschnittliche Krippendorffs-Alpha für die für diese Analyse relevanten Codes für Erfahrung (EXP), Wissen (KNO), Hinweise (ADV), erfahrungsbasierte Lösungsvorschläge (SUG\_EXP), wissensbasierte Lösungsvorschläge (SUG\_KNO) liegt bei .97 (min. .94, max. .98). Für die Analyse wurde nur die Teilmenge der Daten genutzt, bei der es eine perfekte Übereinstimmung gab, um die Datenqualität weiter zu erhöhen. In mehreren logistischen Regressionen wird der Einfluss der Prompts und der vorherigen demografischen Daten auf die jeweiligen reflexionsbezogenen Codes überprüft.

Die erste Hypothese besagt, dass die Erfahrungs-Prompts zu mehr Erfahrungen in den Texten führen als die beiden anderen Prompts. Die Hypothese wird über den Code für Erfahrungen (EXP) operationalisiert. Es wurde eine binär logistische Regression durchgeführt, um zu prüfen, ob die Variable Prompts einen Einfluss auf den Code für Erfahrungen hat. Das Modell ist signifikant für die Prädiktoren Prompt und Zeit ( $\chi^2=71,932$ ,  $df=12$ ,  $p<,001$ ). Die anderen Prädiktoren (Altersgruppe, Geschlecht, wöchentliche Arbeitsstunden, Academics, ExpMoreThanAYear, und NativeSpeaker) fügen nicht signifikant zum Modell hinzu. Die Variable für die Anzahl der Wörter ist zwar für das Gesamtmodell als signifikanter Faktor angegeben, aber wenn man die einzelnen Ausprägungen der Variablen im Modell kontrastiert (SPSS *indicator* contrast), dann ist ersichtlich, dass keine der Ausprägungen der Variable Zeit signifikant ist. Das Modell erklärt 22,8 % (Cox & Snell  $R^2$ ) bzw. 30,6 % (Nagelkerke  $R^2$ ) der Varianz des Codes EXP. Das Modell hat 69,7 % der Fälle vorhergesagt, in denen keine Erfahrung vorhanden war (Code EXP). Es wurden 71,1 % der Fälle vorhergesagt, in denen Erfahrungen im Text vorhanden waren. Somit hat das Modell insgesamt 70,5 % der Erfahrungen (Code EXP) korrekt vorhergesagt. Die Tabelle 5-5 zeigt den Einfluss des Erfahrungs-Prompts (EP) auf den Code EXP im Vergleich zu dem Kontroll-Prompt (KP) und dem Lösungs-Prompt (LP). Tabelle 5-5 zeigt, dass der Erfahrungs-Prompt die Wahrscheinlichkeit für Erfahrungen im Text im Vergleich zu den beiden anderen Prompts um einen Faktor von ca. 8-10 erhöht.

In Tabelle 5-5 ist ebenfalls ersichtlich, dass die Anzahl der Wörter bei dem Code für Erfahrung eine Rolle spielt. Laut dem Vergleich scheint es so, als sei die



Wahrscheinlichkeit, um in einem Text einen Code für Erfahrung vorzufinden um Faktor 3-4 höher, sobald der Text mehr als 80 Wörter aufweist.

**Tabelle 5-5: Einfluss von Prompts und der Anzahl der Wörter auf Erfahrungen (Code EXP)**

Prompt	Vgl.	B (S.E.)	Wald	Sig.	EXP(B)	95 % CI für (EXP(B))	
						Lower	Upper
<b>EP</b>	KP	2,096 (.381)	30,278	<,001	8,131	3,854	17,152
<b>EP</b>	LP	2,358 (.397)	35,250	<,001	10,572	4,853	23,027
<b>W0-40</b>	W121+	1,188 (.550)	4,660	,031	3,281	1,116	9,652
<b>W41-80</b>	W81-120	1,171 (.395)	8,793	,003	3,225	1,487	6,993
<b>W41-80</b>	W121+	1,386 (.496)	7,798	,005	3,997	1,512	10,572

Die zweite Hypothese besagt, dass durch den Lösungs-Prompt mehr allgemeine Lösungen entstehen (also keine erfahrungsbasierten oder wissensbasierten Lösungen; Code SUG) als bei den anderen beiden Prompts. Operationalisiert wird dies über den reflexionsbezogenen Code SUG. Das Modell ist für die Prädiktoren Prompt und Wörter signifikant ( $\chi^2=65,686$ ,  $df=12$ ,  $p<,001$ ). Die anderen Prädiktoren (Zeit, Altersgruppe, Geschlecht, wöchentliche Arbeitsstunden, Academics, ExpMoreThanAYear, und NativeSpeaker) fügen nicht signifikant zum Modell hinzu. Der Faktor NativeSpeaker ist allerdings mit einem  $p=.054$  knapp an der Grenze. Das Modell erklärt 52,8 % der nicht-kodierten Vorschläge und 80,6 % der kodierten Vorschläge, wodurch insgesamt 69,8 % der Lösungsvorschläge durch das Modell korrekt vorhergesagt werden. Das Modell erklärt 21 % (Cox & Snell  $R^2$ ) bzw. 28,5 % (Nagelkerke  $R^2$ ) der Varianz allgemeiner Lösungsvorschläge (SUG). Tabelle 5-6 zeigt, dass durch einen Lösungs-Prompt die Wahrscheinlichkeit auf allgemeine Lösungsvorschläge im Vergleich zu den anderen beiden Prompts ca. 4,8x höher ist.

In Tabelle 5-6 ist ebenfalls ersichtlich, dass die Anzahl der Wörter eine Rolle spielt. Die Ergebnisse zeigen, dass eine sehr hohe Anzahl an Wörtern ( $>120$ ) die Wahrscheinlichkeit auf allgemeine Lösungsvorschläge in einem Text im Vergleich zu sehr wenigen Wörtern (0-40) um den Faktor 21,5 erhöht. Je nach Konstellation sinkt die Wahrscheinlichkeit, aber der Trend ist eindeutig.

Die dritte Hypothese zielt nicht wie die zwei vorherigen Hypothesen darauf ab, dass exakt das passiert, was der Prompt versucht anzuregen, sondern vermutet, dass beim Erfahrungs-Prompt mehr erfahrungsbasierte Lösungsvorschläge entstehen als bei den beiden anderen Prompts. Dementsprechend wird diese Hypothese über den Code für erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) operationalisiert. Das Modell ist für die Prädiktoren Prompt, NativeSpeaker und die Anzahl der Wörter signifikant ( $\chi^2=50,233$ ,  $df=12$ ,  $p<,001$ ). Die anderen Prädiktoren (Zeit, Altersgruppe, Geschlecht,

wöchentliche Arbeitsstunden, Academics, und ExpMoreThanAYear) fügen nicht signifikant zum Modell hinzu. Durch die drei Variablen werden 16,5 % (Cox & Snell  $R^2$ ) bzw. 24 % (Nagelkerke  $R^2$ ) des Modells erklärt. Das Modell erklärt 92,6 % der nicht-kodierten erfahrungsbasierten Lösungsvorschläge und 28 % der kodierten erfahrungsbasierten Lösungsvorschläge, was insgesamt 75,2 % der erfahrungsbasierten Lösungsvorschläge abbildet.

**Tabelle 5-6: Einfluss der Prompts und der Anzahl der Wörter  
auf allgemeine Lösungsvorschläge (Code SUG)**

Prompt	Vgl.	B (S.E.)	Wald	Sig.	EXP(B)	95 % CI für (EXP(B))	
						Lower	Upper
<b>LP</b>	KP	1,582 (.380)	17,351	<,001	4,865	2,311	10,242
<b>LP</b>	EP	1,564 (.374)	17,478	<,001	4,777	2,295	9,944
<b>W0-40</b>	W41-80	1,145 (.357)	10,318	,001	3,143	1,563	6,323
<b>W0-40</b>	W81-120	1,585 (.456)	12,088	,001	4,879	1,997	11,924
<b>W0-40</b>	W121+	3,070 (.709)	18,723	<,001	21,534	5,361	86,491
<b>W41-80</b>	W121+	1,924 (.662)	8,446	,004	6,850	1,871	25,081
<b>W81-120</b>	W121+	1,485 (.706)	4,417	,036	4,413	1,105	17,620

Die Tabelle 5-7 zeigt, dass zwar der Erfahrungs-Prompt (EP) mehr erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) hervorbringt als der Kontroll-Prompt (KP), aber dass dieser im Vergleich zum Lösungs-Prompt (LP) nicht signifikant mehr Lösungsvorschläge hervorbringt. Der Lösungs-Prompt erhöht aber im Vergleich zum Kontroll-Prompt signifikant die Wahrscheinlichkeit erfahrungsbasierter Lösungsvorschläge. So lässt sich feststellen, dass sowohl Erfahrungs-Prompts als auch Lösungs-Prompts die Wahrscheinlichkeit auf Lösungen erhöhen (um Faktor 4,4 bzw. um Faktor 2,3). In diesem Modell hat die Sprache der Crowdworker ebenfalls einen Einfluss und es zeigt sich, dass Personen, die Englisch als Muttersprache sprechen, die Wahrscheinlichkeit auf erfahrungsbasierte Lösungsvorschläge um einen Faktor von 2,9 erhöhen. Der Faktor ist höher als der Faktor bei Einsatz der Lösungs-Prompts. Bei dem Code für erfahrungsbasierte Lösungsvorschläge spielt die Anzahl der Wörter ebenfalls eine starke Rolle. Durch die Zahlen lässt sich ablesen, dass mehr Wörter hilfreich sind, um die Wahrscheinlichkeit auf erfahrungsbasierte Lösungen zu erhöhen. Insbesondere längere Texte mit 81-120 und mehr Wörtern haben eine hohe Wahrscheinlichkeit auf erfahrungsbasierte Lösungen. Die Wahrscheinlichkeit liegt höher als die Wahrscheinlichkeit bei Erfahrungs-Prompts. Um einen möglichen Einfluss der Sprache auf die Anzahl der Wörter auszuschließen, wurde ein separates Modell berechnet. Innerhalb des Datensatzes wurden dafür nur Einträge berücksichtigt, in denen die Personen Englisch als Muttersprache angegeben haben. Im Regressionsmodell wurden alle demografischen Daten bis auf die Sprache berücksichtigt, da dieser Wert nun

konstant ist. Das Modell zeigt, dass auch hier die Wörter einen Einfluss haben: Es wurden Korrelationen von 0-40 auf 41-80 Wörter, 0-40 auf 121+ und 41-80 auf 121+ Wörter gefunden. Für die Personen, die Englisch nicht als Muttersprache sprechen, wurden keine Regressionsmodelle berechnet, da die Anzahl der Datensätze dafür jeweils zu niedrig war (siehe dazu auch Abbildung 5-5).

**Tabelle 5-7: Einfluss der Prompts, der Sprache und der Anzahl der Wörter  
auf erfahrungsbasierte Lösungsvorschläge (Code SUG\_EXP)**

Prompt	Vgl.	B (S.E.)	Wald	Sig.	EXP(B)	95 % CI für (EXP(B))	
						Lower	Upper
<b>EP</b>	KP	1,489 (.393)	14,328	<,001	4,432	2,050	9,580
<b>EP</b>	LP	,621 (.355)	3,058	,080	1,862	,928	3,736
<b>LP</b>	KP	,867 (.411)	4,444	,035	2,380	1,063	5,332
<b>Native Speaker</b>	Non-Native	1,092 (.360)	9,208	,002	2,981	1,472	6,038
<b>W0-40</b>	W41-80	1,249 (.500)	6,244	,012	3,486	1,309	9,281
<b>W0-40</b>	W81-120	1,829 (.559)	10,693	,001	6,229	2,081	18,646
<b>W0-40</b>	W121+	2,506 (.613)	16,723	<,001	12,260	3,688	40,756
<b>W41-80</b>	W121+	1,258 (.454)	7,658	,006	3,517	1,443	8,572

Die vierte Hypothese beinhaltet, dass der Kontroll-Prompt mehr unbegründete Hinweise nach sich zieht als die beiden anderen Prompts. Operationalisiert wird dies über den Code für Hinweise (ADV). Das Modell ist signifikant für die Prädiktoren Prompt und NativeSpeaker ( $\chi^2=24,158$ ,  $df=10$ ,  $p=,007$ ). Die übrigen Prädiktoren (Zeit, Wörter, Altersgruppe, Geschlecht, wöchentliche Arbeitsstunden, Academics, und ExpMoreThanAYear) tragen nicht signifikant zum Modell bei. Durch die Prompts, die Anzahl der Wörter und die Variable NativeSpeaker werden 8,3 % (Cox & Snell  $R^2$ ) bzw. 11,6 % (Nagelkerke  $R^2$ ) des Modells erklärt. Das Modell erklärt 93,5 % der nicht-kodierten Hinweise und 21,7 % der kodierten Hinweise, was insgesamt 69,8 % der Hinweise korrekt vorhersagt.

Aus Tabelle 5-8 geht hervor, dass der Kontroll-Prompt (KP) im Vergleich zum Lösungs-Prompt (LP) nicht signifikant mehr Vorschläge hervorruft. Allerdings ist der Unterschied vom Kontroll-Prompt zum Erfahrungs-Prompt (EP) signifikant. Der Lösungs-Prompt erzeugt signifikant mehr Hinweise als der Erfahrungs-Prompt. Obwohl die Sprache als Variable im Modell signifikant war, zeigt sich in diesem Modell kein signifikanter Einfluss derjenigen, die Englisch als Muttersprache sprechen.

Tabelle 5-8: Einfluss der Prompts und der Sprache auf Hinweise (Code ADV)

Prompt	Vgl.	B (S.E.)	Wald	Sig.	EXP(B)	95 % CI für (EXP(B))	
						Lower	Upper
<b>KP</b>	LP	,419 (.318)	1,738	,187	1,521	,815	2,837
<b>KP</b>	EP	1,283 (.347)	13,647	<,001	3,608	1,826	7,127
<b>LP</b>	EP	,864 (.335)	5,925	,015	2,372	1,183	7,127
<b>Native Speaker</b>	Non-Native	,578 (.306)	3,580	,058	1,783	,979	3,247

Da bisher sowohl allgemeine (SUG) und erfahrungsbasierte (SUG\_EXP) Lösungsvorschläge analysiert wurden, erfolgt noch eine Analyse der wissensbasierten Lösungsvorschläge (SUG\_KNO).

Das Modell ist signifikant für die Prädiktoren Prompt und Anzahl der Wörter ( $\chi^2=57,594$ ,  $df=12$ ,  $p<,001$ ). Die übrigen Prädiktoren (Zeit, Altersgruppe, Geschlecht, wöchentliche Arbeitsstunden, Academics, NativeSpeaker, und ExpMoreThanAYear) tragen nicht signifikant zum Modell bei. Durch die Prompts und die Anzahl der Wörter werden 18,7 % (Cox & Snell  $R^2$ ) bzw. 25,3 % (Nagelkerke  $R^2$ ) des Modells erklärt. Das Modell erklärt 93,1 % der nicht-kodierten wissensbasierten Lösungsvorschläge und 55,4 % der kodierten wissensbasierten Lösungsvorschläge, wodurch das Modell insgesamt 71,9 % der wissensbasierten Lösungsvorschläge korrekt vorhersagt.

Tabelle 5-9: Einfluss der Prompts auf wissensbasierte Lösungsvorschläge (Code SUG\_KNO)

Prompt	Vgl.	B (S.E.)	Wald	Sig.	EXP(B)	95 % CI für (EXP(B))	
						Lower	Upper
<b>LP</b>	KP	,796 (.323)	6,059	,014	2,216	1,176	4,174
<b>KP</b>	EP	1,459 (.369)	15,671	<,001	4,301	2,089	8,856
<b>LP</b>	EP	2,254 (.374)	36,289	<,001	9,529	4,576	19,843
<b>W0-40</b>	W41-80	,808 (.381)	4,492	,034	2,242	1,063	4,732
<b>W0-40</b>	W81-120	1,043 (.456)	5,224	,022	2,837	1,160	6,939
<b>W0-40</b>	W121+	1,692 (.540)	9,828	,002	5,433	1,886	15,651

Die Tabelle 5-9 zeigt, dass der Lösungs-Prompt (LP) signifikant mehr wissensbasierte Lösungsvorschläge (SUG\_KNO) erzeugt als der Kontroll-Prompt (KP) und der Erfahrungs-Prompt (EP). Der Kontroll-Prompt erzeugt ebenfalls mehr wissensbasierte Lösungsvorschläge als der Erfahrungs-Prompt. Im Vergleich zum Erfahrungs-Prompt erhöht der Lösungs-Prompt die Wahrscheinlichkeit für einen wissensbasierten Lösungsvorschlag um einen Faktor von 9,5. Die Anzahl der Wörter spielt ebenfalls eine Rolle. Die Daten zeigen, dass der Text deutlich wahrscheinlicher einen Lösungsvorschlag enthält, wenn er zumindest mehr als 40 Wörter enthält. Von den mittleren Gruppen der Anzahl der Wörter (bspw. 41-80 und 81-120) gab es keine signifikanten Unterschiede zu sehr vielen Wörtern (121+).

### 5.1.6 Diskussion

In diesem Abschnitt werden die Ergebnisse kritisch diskutiert. Zunächst werden die in Abschnitt 5.1.3 hergeleiteten Hypothesen überprüft, über die auch die Forschungsfrage adressiert wird. Anschließend werden die übrigen Ergebnisse interpretiert. Es folgt eine Ableitung von Designempfehlungen für Prompts basierend auf den Ergebnissen.

#### 5.1.6.1 Überprüfung der Hypothesen

Hypothese 1 vermutet, dass mehr Erfahrungen durch den Erfahrung-Prompt (EP) als durch die beiden anderen Prompts (Z6) angeregt werden. Die Regressionsanalyse hat signifikante Effekte gefunden, die zeigen, dass der Erfahrung-Prompt im Vergleich zu dem Kontroll-Prompt (KP) und dem Lösungs-Prompt (LP) häufiger zu Erfahrungen (Code EXP) in den Texten führt (siehe Tabelle 5-5). Zwischen dem Kontroll-Prompt und dem Lösungs-Prompt wurden keine signifikanten Unterschiede in Bezug auf geschriebene Erfahrung gefunden. Somit zeigt sich:  $EP(EXP) > LP(EXP), KP(EXP)$ . Dies entspricht der Gleichung, die für die Hypothese initial aufgestellt wurde. Somit kann Hypothese 1 gehalten werden.

In Hypothese 2 ist die Vermutung, dass der Lösungs-Prompt (LP) mehr Lösungen produziert als die beiden anderen Prompts (Z8). Eine Regressionsanalyse zeigt, dass der Lösungs-Prompt signifikant mehr allgemeine Lösungsvorschläge (SUG) als der Kontroll-Prompt (KP) und Erfahrungs-Prompt (EP) hervorbringt (siehe Tabelle 5-6). Zwischen Kontroll-Prompt und Erfahrungs-Prompt gibt es hinsichtlich allgemeiner Lösungsvorschläge (SUG) keine Unterschiede. Dementsprechend ist die Gleichung:  $LP(SUG) > EP(SUG), KP(SUG)$ . Dies entspricht der Hypothese und somit kann Hypothese 2 beibehalten werden.

Die Daten zeigen, dass die beiden Hypothesen 1 und 2 behalten werden können. Dies zeigt, dass Prompts in dieser Teilstudie Elemente, wie Erfahrungen und Lösungsvorschläge, gut anregen können. Dementsprechend können die beiden Ziele Z6 und Z8 durch Prompts angeregt werden.

Die dritte Hypothese vermutet, dass durch den Erfahrungs-Prompt beeinflusste Antworten eher erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) enthalten als die Antworten, die durch den Lösung-Prompt (LP) oder Kontroll-Prompt (KP) beeinflusst werden. Die Regressionsanalyse zeigt, dass der Erfahrungs-Prompt signifikant mehr erfahrungsbasierte Lösungen erzeugt als der Kontroll-Prompt (siehe Tabelle 5-7). Der Unterschied zwischen dem Erfahrungs-Prompt und dem Lösungs-Prompt ist nicht

signifikant. Der Lösungs-Prompt erzeugt ebenfalls mehr erfahrungsbasierte Lösungsvorschläge als der Kontroll-Prompt. Somit ist die Gleichung:  $EP(SUG\_EXP), LP(SUG\_EXP) > KP(SUG\_EXP)$ . Dies entspricht nicht der Hypothese und somit muss Hypothese 3 verworfen werden. Laut dem Ergebnis eignen sich wohl der Erfahrungs- als auch der Lösungs-Prompt gleichermaßen, um erfahrungsbasierte Lösungsvorschläge anzuregen. Dies impliziert, dass in dieser Teilstudie die Teilnehmer bei Lösungsvorschlägen häufig eigene Erfahrungen eingebracht haben.

Hypothese 4 vermutet, dass der Kontroll-Prompt (KP) im Vergleich zu den beiden anderen Prompts häufiger zu unbegründeten Hinweisen (ADV) führt. Die Regressionsanalyse zeigt, dass der Kontroll-Prompt signifikant mehr Hinweise als der Erfahrungs-Prompt erzeugt (siehe Tabelle 5-8). Der Unterschied zum Lösungs-Prompt ist nicht signifikant. Somit ist die Gleichung:  $KP(ADV), LP(ADV) > EP(ADV)$ . Dies entspricht nicht der Vermutung in der Hypothese. Dementsprechend wird Hypothese 4 verworfen.

#### **5.1.6.2 Interpretation der übrigen Variablen**

Die Regressionsanalyse für wissensbasierte Lösungsvorschläge (SUG\_KNO) zeigt, dass der Lösungs-Prompt (LP) signifikant häufiger Lösungsvorschläge erzeugt als die beiden anderen Prompts (siehe Tabelle 5-9). Der Kontroll-Prompt (KP) erhöht im Vergleich zum Erfahrungs-Prompt (EP) ebenfalls die Wahrscheinlichkeit auf wissensbasierte Lösungsvorschläge. Somit lässt sich für wissensbasierte Lösungsvorschläge die folgende Gleichung aufstellen:  $LP(SUG\_KNO) > KP(SUG\_KNO) > EP(SUG\_KNO)$ .

Es lässt sich feststellen, dass der Lösungs-Prompt bei vielen reflexionsbezogenen Codes einen signifikanten Einfluss hat. Der Lösungs-Prompt produziert allgemein mehr Lösungsvorschläge (SUG) und mehr wissensbasierte Lösungsvorschläge (SUG\_KNO) als die beiden anderen Prompts. Der Prompt produziert zusammen mit dem Erfahrungs-Prompt mehr erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) als der Kontroll-Prompt. Der Lösungs-Prompt produziert ebenfalls mehr Hinweise (ADV) als der Kontroll-Prompt. Hier stellt sich die Frage, ob es möglich ist, die Formulierung des Lösungs-Prompts weiter zu schärfen, um die Wirkung besser auf einzelne Elemente von Reflexion (sprich einzelne Codes) fokussieren zu können. Da (kollaborative) Reflexion auf dem Austausch von Erfahrungen basiert (siehe Abschnitt 2.5), ist das Ziel, den Lösungs-Prompt eher auf erfahrungsbasierte Lösungsvorschläge auszurichten anstatt auf wissensbasierte Lösungsvorschläge oder Hinweise. Hier sind aber weitere Studien notwendig, um zu prüfen, inwieweit dies mit Prompts gelingen kann.

Bei allen untersuchten reflexionsbezogenen Codes hat die Anzahl der Wörter in den Modellen der Regressionsanalyse eine Rolle gespielt. Bei allen Codes gab es einen signifikanten Unterschied zwischen sehr wenigen Wörtern (0-40) und sehr vielen Wörtern (121+), wobei die höhere Anzahl der Wörter immer mit einer höheren Wahrscheinlichkeit verbunden war, den Code im Text vorzufinden. Allgemein lässt sich zusammenfassen, dass bei den meisten reflexionsbezogenen Codes eine höhere Wahrscheinlichkeit für das Auftreten eines Codes vorhanden ist, je mehr Wörter die Verfasser schreiben. Da es oft keinen signifikanten Unterschied zwischen vielen Wörtern (81-120) und sehr vielen Wörtern (121+) lässt sich vermuten, dass es eine Art Grenznutzen gibt und bereits geschätzte 80+ Wörter ausreichen könnten. Dies impliziert, dass sehr wenige Wörter nicht ausreichen, um eine Idee, einen Hinweis oder einen Lösungsvorschlag (sei es allgemein, auf Wissensbasis oder auf Erfahrungsbasis) zu schreiben.

Der Faktor der englischen Sprache ist schwierig zu interpretieren. Es gibt bei erfahrungsbasierten Lösungsvorschlägen (SUG\_EXP) einen signifikanten Einfluss derjenigen, die Englisch als Muttersprache sprechen, aber nicht bei Erfahrungen (EXP) oder allgemeinen (SUG) bzw. wissensbasierten (SUG\_KNO) Lösungsvorschlägen. Eine Interpretation ist, dass es den Personen leichter fällt, aus den eigenen Erfahrungen heraus einen Lösungsvorschlag zu formulieren als Personen, die nicht Englisch als Muttersprache sprechen. Ebenfalls hat die Variable NativeSpeaker einen Einfluss auf Hinweise (ADV). Hier ist aber zu vermuten, dass dies an typischen Floskeln liegt. Sätze, die oft am Ende einer Antwort geschrieben werden, wie bspw. *„Don't give up, just try other strategies to gain success“* oder *„Don't let yourself down mate.“* beinhalten einen Vorschlag und werden deswegen dementsprechend kodiert. Diese Sätze können allerdings als Floskel interpretiert werden und daher kann dieser beobachtete Einfluss daher stammen, dass die Personen, die Englisch als Muttersprache sprechen, wesentlich eher dazu neigen, solche Floskeln am Ende eines Textes anzufügen.

Allgemein lässt sich feststellen, dass die reflexionsbezogenen Codes von den Prompts, und teilweise von der Anzahl der Wörter oder von den Fähigkeiten mit der englischen Sprache der Crowdworker beeinflusst wurden. Es hat sich gezeigt, dass das Alter, das Geschlecht, die Erfahrung als Crowdworker (mehr oder weniger als ein Jahr), die wöchentlichen Arbeitsstunden oder der Bildungsgrad (Akademiker oder Nicht-Akademiker) keinen Einfluss auf die reflexionsbezogenen Codes haben. Das zeigt, dass die Prompts breit anwendbar sind und somit in diesen Modellen keine Anpassung der Prompts auf den Nutzer selbst notwendig sind.

### **5.1.6.3 Möglichkeiten für das Design von Prompts**

Aus den Ergebnissen ergeben sich unterschiedliche Aspekte, die ein Design einer Software zur Unterstützung von online kollaborativer Reflexion beeinflussen. Durch das Beibehalten von Hypothese 1 und 2, lässt sich ableiten, dass Prompts allgemein in Online-Foren bestimmte Themen anregen könnten. In diesem Datensatz scheinen Prompts sich erfolgreich zum Anregen von Diskussionen und Reflexion zu eignen. Die Daten zeigen, dass mit den Prompts für Erfahrungen (EP) und Lösungsvorschläge (LP) beide Aspekte direkt angeregt werden können. Der Prompt für Erfahrung kann zur Anregung von Reflexion genutzt werden. Bei dem Prompt für Lösungsvorschläge muss beachtet werden, dass dieser zusätzlich unbegründete Hinweise und wissensbasierte Lösungsvorschläge anregt, die nicht zwingend förderlich für Reflexion sein müssen.

Gleichzeitig kann man mit dem Lösungs- und dem Erfahrungs-Prompt versuchen, die Konversation an unterschiedlichen Zeitpunkten unterschiedlich zu steuern. Ein System kann beispielsweise zunächst Erfahrungs-Prompts zeigen, um den Erfahrungsaustausch zu fördern und dann nach einigen Antworten Lösungs-Prompts anzeigen, um die kollaborative Reflexion in eine Phase zu bewegen, in der künftige Handlungen geplant werden.

Da die Anzahl der Wörter einen Einfluss auf die reflexionsbezogenen Codes hat, scheint dies ein Faktor zu sein, der in der Praxis wichtig ist. Aus den Ergebnissen ergibt sich die Frage, ob es mit Prompts möglich wäre, die Länge von Beiträgen direkt zu beeinflussen. Beispielsweise kann ein System erkennen, dass es ein sehr kurzer Beitrag ist und dazu passend einen Prompt anbieten, der dem Nutzer vorschlägt, mehr Informationen einfließen zu lassen, damit die Chance steigt, dass der Text für den Empfänger nützlich ist.

### **5.1.7 Einschränkungen**

Es muss unterstrichen werden, dass die komplette Situation für die Probanden lediglich simuliert wurde. Zusätzlich wurde den Teilnehmern anfangs erklärt, dass ein Kommentar zu einem Text abgegeben werden muss. Das kann die Ergebnisse abschwächen. Es wurde versucht, durch Auswahl eines für Crowdworker bekanntes Problem, eine realistische Umgebung zu simulieren. Zudem zeigen die Ergebnisse, dass die Daten hinsichtlich der Demografie balanciert sind. Gleichzeitig hat die überwiegende Mehrheit sinnvolle Kommentare abgegeben und es wurden nur sehr wenige Kommentare aussortiert. Dies zeigt, dass die Crowdworker die Umgebung ernst genommen haben. Diese Faktoren stärken die Ergebnisse.



Zunächst lässt sich feststellen, dass sich in dem simulierten Forum lediglich ein Beitrag befand. Das bedeutet, dass die dargestellten Ergebnisse nur in diesem Fall gültig sind und es sein kann, dass sich die Prompts bei vorhandenen Antworten in einem Thema anders verhalten. So ist unklar, ob sich bspw. Prompts für Lösungsvorschläge anders verhalten, wenn bereits eine Diskussion mit mehreren Antworten vorhanden ist. Die Prompts in diesem Versuchen waren ebenfalls stark unterschiedlich ausgerichtet, was es vermutlich den Probanden erleichtert hat, unterschiedlich darauf zu reagieren. Inwieweit sich Prompts eignen, um Reflexion anzuregen, wenn die Formulierungen dichter beieinanderliegen, hat diese Teilstudie nicht abgedeckt. Dementsprechend ist die Aussagekraft dieser Studie auf das abgebildete Szenario limitiert und die Ergebnisse können nicht für alle Diskussionssituationen generalisiert werden. Die Wirkung von Formulierungen, die dichter beieinanderliegen und die Beantwortung von Themen mit Antworten wird in Teilstudie 2.2 behandelt (siehe Abschnitt 5.2).

In einer Crowdsourcing-Studie lassen sich externe Faktoren nicht kontrollieren (bspw. Ablenkungen durch TV oder Smartphone). Daher scheinen die Ergebnisse eingeschränkt. Auf der anderen Seite sind Arbeitsplätze bzw. die heimische Umgebung bei Nutzung von Foren ebenfalls durch Ablenkungen geprägt. Dementsprechend ist in diesem Versuch das vermeintliche Problem, die externen Faktoren nicht kontrollieren zu können, keine Herausforderung und womöglich eher als positiv einzuschätzen.

Es zeigt sich in den Regressionsmodellen ebenfalls, dass die vorhandenen Variablen zwischen 11 % und 30,6 % (Nagelkerke  $R^2$ ) der Ausprägungen der Codes erklären konnten. Dies zeigt deutlich, dass in allen Fällen noch Faktoren existieren, die nicht identifiziert sind und einen großen Einfluss auf das Vorhandensein der reflexionsbezogenen Codes haben.

Mit der binär logistischen Regression lässt sich der Einfluss mehrerer unabhängiger Variablen auf eine abhängige Variable prüfen. In dieser Teilstudie gab es mit den reflexionsbezogenen Codes allerdings mehrere abhängige Variablen. Diese wurden in mehreren einzelnen binär logistischen Regressionen unabhängig voneinander geprüft. Obwohl die unabhängigen Variablen keine Multikollinearität aufweisen, ist nicht auszuschließen, dass die unterschiedlichen Modelle abhängig voneinander sind.

Die genutzten Prompts decken nur zwei Aspekte von Reflexion ab (Erfahrung und Lösungsvorschläge). Somit sind weitere Studien notwendig, um zu prüfen, inwieweit Prompts zusätzlich andere Aspekte von Reflexion unterstützen können.

### **5.1.8 Zusammenfassung**

Diese Teilstudie hat gezeigt, dass es mit Prompts möglich ist, sowohl Erfahrungen (Z6) als auch Lösungsvorschläge (Z8) in einem simulierten Forum direkt anzuregen. Somit kann Reflexion bzw. können diese entsprechenden Teile der Reflexion direkt angeregt und unterstützt werden.

Beim praktischen Einsatz der Prompts ist zu beachten, dass der Lösungs-Prompt nicht nur erfahrungsbasierte Lösungen (SUG\_EXP) stimuliert, sondern auch wissensbasierte Lösungen (SUG\_KNO) und unbegründete Hinweise (ADV) nach sich zieht.

Ein großer Faktor beim Einsatz der Prompts war die Anzahl der Wörter und es zeigt sich bei allen Codes, dass mehr Wörter die Wahrscheinlichkeit auf das Vorhandensein des entsprechenden reflexionsbezogenen Codes erhöhen. Darüber ergeben sich Möglichkeiten künftiger Unterstützung, indem versucht wird, dass Prompts direkt die Anzahl der Wörter zu beeinflussen, falls jemand im Begriff ist in einer Online-Diskussion einen sehr kurzen Beitrag einzubringen. Darüber könnte sich ebenfalls Reflexion unterstützen lassen. Dies ist in künftigen Studien zu überprüfen.

## **5.2 Teilstudie 2.2: Spezifität von Prompts im Bereich von Lösungs-Prompts**

Im Anschluss an die vorherige Teilstudie, werden in dieser Teilstudie die Auswirkungen einer vorhandenen Diskussion im Vergleich zu einem frischen Thema ohne Antworten auf das Generieren von Lösungsvorschlägen überprüft. Durch leichte Abwandlungen der Prompts wird ebenfalls überprüft, welche Unterschiede leichte Formulierungsänderungen der Prompts haben. Diese Studie wurde ca. 1 Monat nach der vorherigen Teilstudie durchgeführt. Eine oberflächliche Beschreibung der Durchführung findet sich auch in (EmployID 2016, 31–35).

### **5.2.1 Ansatz**

Die vorherige Teilstudie hat gezeigt, dass sich bestimmte Elemente von Reflexion, wie Erfahrungen oder Lösungsvorschläge in einer simulierten Forenumgebung direkt durch Prompts anregen lassen. Im Gegensatz zur vorherigen Teilstudie, wird in dieser Teilstudie ein Forum in einer Crowdsourcing-Umgebung simuliert, in der bereits Diskussionsbeiträge enthalten sind. Auf diese Weise soll analysiert werden, inwieweit das Vorhandensein von Diskussionsbeiträgen die Wirkung von Prompts beeinflusst.

Für das Vorhandensein von Lösungsvorschlägen in der Diskussion wird die simulierte Forenumgebung um drei Antworten erweitert, die alle auf Antworten von Crowdworkern aus Teilstudie 2.1 basieren (siehe Abschnitt 5.1). Probanden in dieser Teilstudie sollten eine Antwort auf die fiktive Problemstellung eines Crowdworkers geben.

Aus der vorherigen Teilstudie ist ersichtlich geworden, dass sehr deutlich unterschiedlich formulierte Prompts einen Einfluss auf die Antworten haben. Ob die Unterschiede in den reflexionsbezogenen Codes bestehen bleiben, wenn sich die Prompts weniger stark unterscheiden, ist nicht eindeutig. In dieser Teilstudie werden drei Prompts untersucht, die alle thematisch Lösungsvorschläge zum Inhalt haben (Z8), sich aber in der Formulierung leicht unterscheiden. Es soll evaluiert werden, *inwieweit feinere Nuancen in den Formulierungen von Prompts eine Rolle beim Anregen von Lösungsvorschlägen für kollaborative Reflexion spielen.*

Zusätzlich wird der Lösungs-Prompt aus der vorherigen Teilstudie auch in dieser Teilstudie verwendet. Das Ziel ist ebenfalls zu prüfen, *inwieweit das Vorhandensein von Diskussionsbeiträgen einen Einfluss auf die Erstellung von Lösungsvorschlägen in einer simulierten Community of Practice Plattform hat.*

## **5.2.2 Methode**

In diesem Abschnitt wird die Methodik der Studie erklärt. Dazu wird ein Überblick über die Prompts, den Versuchsaufbau und der Konfiguration in Crowdfunder gegeben und das geplante Vorgehen für die Analyse erläutert.

### **5.2.2.1 Umsetzung der Teilstudie**

In dieser Teilstudie werden insgesamt drei Prompts miteinander verglichen (siehe Übersicht in Tabelle 5-10). Da analysiert werden soll, welchen Einfluss das Vorhandensein existierender Lösungsvorschläge auf neue Antworten hat, wird der Lösungs-Prompt aus der vorherigen Teilstudie direkt übernommen (siehe Tabelle 5-1). Auf diese Weise ist der einzige Unterschied zwischen den beiden Studien das Vorhandensein (in dieser Teilstudie) bzw. die Abwesenheit (in Teilstudie 2.1) von existierenden Lösungsvorschlägen.

Tabelle 5-10: Prompts in Teilstudie 2.2

Name	Abkürzung	Prompt
<b>Lösungs-Prompt</b>	LP	<b>Enter a comment.</b> Describe a solution to the problem above and why you think this is a solution.
<b>Spezifischer Lösungs-Prompt</b>	SLP	<b>Enter a comment.</b> Which of the solutions would you like to take and why?
<b>Kontroll-Lösungs-Prompt</b>	KLP	<b>Enter a comment.</b> Refer to the other comments.

Zusätzlich soll geprüft werden, ob eine leichte Änderung in der Formulierung eine Auswirkung auf die reflexionsbezogenen Codes hat. Dazu wird ein Prompt genutzt, der danach fragt, welche anderen Lösungen der Proband bevorzugen würde und warum. Dieser Prompt bezieht sich auf den kollaborativen Aspekt in der kollaborativen Reflexion, bei dem aktiv Ideen diskutiert werden (Dyke 2006; Nicholson u. a. 2003; Woerkom und Croon 2008; Kim und Lee 2002). Der Kontroll-Prompt in dieser Teilstudie unterscheidet sich von dem in der letzten Teilstudie und wird daher als Kontroll-Lösungs-Prompt bezeichnet. Dieser bittet den Empfänger, darum sich explizit auf einen der anderen Kommentare zu beziehen. Somit fokussiert sich diese Teilstudie sowohl auf den Aspekt des Vorhandenseins existierender Lösungen als auch auf Prompts zum Thema Lösungsvorschläge (Z8), die sich inhaltlich nicht sehr stark unterscheiden.

Diese Teilstudie wird in einem between-subjects-Design durchgeführt, was bedingt, dass jeder Proband nur ein einziges Mal teilnimmt. Dementsprechend sieht jeder Proband nur einen einzigen Prompt. Dies hilft Redundanzen und Übertragungseffekte zu vermeiden.

### 5.2.2.2 Abbildung in Crowdfower

Diese Teilstudie wurde in Crowdfower (heute FigureEight) durchgeführt. In diesem Abschnitt werden Maßnahmen zur Qualitätssicherung und die Implementierung der Teilstudie beschrieben.

Diese Teilstudie wurde mit 300 Personen durchgeführt, die zufällig auf die einzelnen Konditionen, sprich die einzelnen Prompts, verteilt wurden. Die ersten 300 Personen, die den Job bei Crowdfower bearbeitet haben, wurden in die Teilstudie aufgenommen. Um das between-subjects-Design in Crowdfower abzubilden, durfte jeder Crowdworker nur ein einziges Mal an der Teilstudie teilnehmen.

Für die Nutzung von Crowdsourcing-Umgebungen in wissenschaftlichen Studien ergeben sich unterschiedliche Anforderungen, die in den Abschnitten 3.6.7 und 3.6.9 beschrieben sind.

Für die Teilstudie werden nur Crowdworker ausgewählt, die intern den höchsten der drei Ränge erreicht haben. Dies bedeutet, dass der Job in der Bearbeitung länger dauert, da es weniger Personen gibt, die diesen Rang haben. Durch den höheren Rang ist die Qualität der Antworten höher und es sinkt die Wahrscheinlichkeit, viele Antworten von Nutzern mit Täuschungsabsicht zu bekommen.

Nachteilig kann die fehlende Kontrollierbarkeit von externen Einflüssen, wie bspw. Geräusche durch einen TV, bei der Durchführung von wissenschaftlichen Studien in einer Crowdsourcing-Umgebung sein. Dies ist in dieser Teilstudie weniger ein Nachteil, da die simulierte Forenumgebung so leicht realistischer ist. Ein regulärer Arbeitsplatz, an dem online kollaborativ reflektiert wird, ist ebenfalls von Ablenkungen geprägt und ähnelt keiner ruhigen Laborumgebung.

Es wurde eine Gold-Standard-Frage implementiert (siehe Abschnitt 3.6.7), die identisch zu der aus der vorherigen Teilstudie ist (siehe Abschnitt 5.1.2.2.1). Sie fragt, wie oft der Autor des fiktiven Problems dieses Problem bereits hatte (zwei Mal). Darüber kann zumindest identifiziert werden, ob der jeweilige Proband das Problem gelesen hat. Dies hilft, leicht Probanden zu identifizieren, die eine Täuschungsabsicht haben und um diese ggf. aus der Teilstudie auszuschließen.

Zur fairen Bezahlung wurde ein Vorabtest durchgeführt, bei dem analysiert wurde, wie lange Probanden für die Teilstudie gebraucht haben. Es wurde ein Stundenlohn von acht US-Dollar pro Stunde anvisiert (Durchführung der Teilstudie in Dezember 2015). Pro Teilnahme wurden 0,80 \$ ausgezahlt und sehr gute Antworten wurden mit zusätzlichen 0,20 \$ prämiert.

Zur Vermeidung von Sprachbarrieren und zur Sicherung der Qualität wurde die Studie auf Teilnehmer aus den folgenden Ländern limitiert (siehe auch Abschnitt 5.1.2.2.2): USA, Großbritannien, Deutschland, Polen, Kroatien, Norwegen, Niederlande, Schweden, Slowenien, Belgien, Österreich, Finnland, Estland und Dänemark.

Die Abbildung 5-6 zeigt die simulierte Forenumgebung in Crowdfunder. In der Abbildung sind die Instruktionen für die Teilstudie nicht enthalten, welche aber in dieser Teilstudie für die Teilnehmer präsent waren. Das fiktive geschilderte Problem wurde aus Teilstudie 2.1 (siehe Abschnitt 5.1.2.2.1) unverändert übernommen. Zusätzlich wurden drei Antworten aufgenommen, die alle einen Beitrag zu der Diskussion leisten. Alle drei

Antworten basieren auf Antworten von Crowdworkern aus Teilstudie 2.1. Die Antworten wurden teilweise leicht angepasst (bspw. abschweifende Sätze entfernt) und sprachlich korrigiert (bspw. Korrektur von Tippfehlern). Das Ziel dieses Vorgehens ist, dass das Szenario für die Crowdworker, die als Probanden an der Teilstudie teilnehmen, möglichst realistisch ist. Jeder Teilnehmer hat dasselbe Szenario und die gleichen Diskussionsbeiträge gesehen. Innerhalb der Studie gab es daher nur hinsichtlich der angezeigten Prompts Unterschiede.

**Original Post:**

Sometimes I get negative feedback on jobs I completed. Then the requester rejects my work and refuses payment although I did my best to answer the questions correctly and truthfully. This happened to me twice last week. Afterwards I get problems motivating myself because I invested so much time and effort. So I sometimes ask myself why should I take more jobs if I always get such negative experiences.

**Comment #1:**

I understand and sympathise with your problem as I have had similar experiences myself. I think the important thing is to remain objective and polite, whilst at the same time contesting the outcome, providing evidence if possible. Often a constructive reply to the requester, explaining that you answered the questions as good as you could and pointing out your high level of consistency in the past will swing things in your favour. If several people have had the same problem with a task it can sometimes result in your reinstatement and/or payment.

**Comment #2:**

Sometimes, you aren't a good match for the task you tried to complete. Even if you tried your best, you sometimes may not have the appropriate skill set to complete the task to the requester's liking. I would recommend exploring different tasks until you find some that match well with your knowledge and skill set. For example, I find that I'm good at the image-related tasks and audio transcriptions, so I jump on those whenever they appear.

**Comment #3:**

To avoid getting overly frustrated, I would suggest taking a break from jobs that require in depth feedback, and do some jobs that are more simplistic and straightforward, just so you can re-energize yourself for a few minutes. Then, when going back to the more complex jobs that require feedback, make sure to read the directions thoroughly, and read any examples of what they are looking for. If you still feel as though you are being rejected unfairly and that you have done the job as they had asked, I would recommend contacting Crowdfunder support about your problem, perhaps the requester of the job is in the wrong.

**Abbildung 5-6: Simulierte Diskussion in Teilstudie 2.1 in Crowdfunder**

Unterhalb der Beiträge in Abbildung 5-6 befand sich, ähnlich wie in Abbildung 5-1, der Prompt und die Textbox zum Verfassen der Antwort. Darunter befanden sich die dieselben demografischen Fragen wie in Teilstudie 2.1. Diese Fragen beinhalten Geschlecht, Alter, Bildungsgrad, wöchentliche Arbeitsstunden als Crowdworker und Vorerfahrung als Crowdworker.

### 5.2.2.3 Geplante Methodik für die Analyse

Zur Überprüfung, ob die Prompts eine Auswirkung auf die Antworten haben, müssen diese inhaltlich untersucht werden. Dazu dient abermals das reflexionsbezogene Kodierschema aus Abschnitt 3.4.

Dieses Kodierschema wurde für diesen Versuch erweitert, um die Antworten für die Prompts SLP und KLP korrekt erfassen zu können (siehe Tabelle 5-11). Bei beiden Prompts wird der Empfänger aufgefordert, sich eine der Antworten auszusuchen und für diese Antwort zu argumentieren bzw. generell eine der Antworten zu referenzieren. Die neuen Codes bilden nicht nur ein „Bezug vorhanden“ ab, sondern zeigen konkret, welchen Inhalt die referenzierte Antwort hat. Zur Erstellung der Codes wurden daher die neu aufgenommenen Antworten in die Diskussion (siehe Abbildung 5-6) analysiert.

Die Kodierung lief analog zu dem Prozess von Teilstudie 2.1 ab (siehe Abschnitt 5.1.2.4).

**Tabelle 5-11: Neue Codes für die inhaltliche Analyse. Ein Prompt schlägt die Nutzung von Vorschlägen aus den vordefinierten Antworten vor. In der Spalte „Kommentar“ ist abgebildet, welcher Vorschlag, in welchem Kommentar unterbreitet wird.**

<b>Abkürzung</b>	<b>Code</b>	<b>Kommentar</b>
A_SUPPORT	Contact Crowdfunder support	#3
A_READ	Read instructions carefully	#3
A_BREAK	Take a break	#3
A_DIFFERENT	No good match for task, find different tasks	#2
A_REQUESTER	Contacting the requester, protesting	#1

### 5.2.3 Hypothesen

Die Grundannahme der Teilstudie ist, dass Prompts eine Auswirkung auf die Antworten der Probanden haben können (siehe 2.10). Zur besseren Analyse der Forschungsfrage in dieser Teilstudie wurden unterschiedliche Hypothesen entwickelt.

Kollaborative Reflexion wird in der Literatur als Prozess in einer Gruppe beschrieben, der durch Erfahrungsaustausch, Diskussion, und dem Hinterfragen von Annahmen geprägt ist, sowie dem Generieren, Diskutieren und Verwerfen von Ideen (Daudelin 1996; Høyrup 2004; Woerkom und Croon 2008; Raelin 2002; Dyke 2006; Hatton und Smith 1995). Es kann daher bspw. leichter sein, einen konstruktiven Lösungsvorschlag einzubringen, wenn bereits andere Lösungsvorschläge vorhanden sind. Dementsprechend ist die zweite Grundannahme dieser Teilstudie, dass das Vorhandensein einer Diskussion einen Unterschied bei der Entwicklung von Lösungsvorschlägen macht. Die Annahme ist, dass ohne vorherige Antworten, weniger neue Lösungsvorschläge eingebracht werden als bei vorhandenen Lösungsvorschlägen.

*Hypothese 1: Der Lösungs-Prompt führt zu mehr Lösungsvorschlägen bei vorhandener Diskussion als in einem frisch gestarteten Thema.*

Diese Hypothese kann beibehalten werden, indem die dazugehörige Nullhypothese widerlegt wird. Operationalisiert wird diese Hypothese über die Menge von Lösungsvorschlägen (SUG). Diese Variable zeigt, ob entweder ein erfahrungsbasierter oder ein wissensbasierter Lösungsvorschlag im Text enthalten ist. Es wird die Menge der Lösungsvorschläge aus dieser Teilstudie mit der Menge der Lösungsvorschläge aus der vorherigen Teilstudie verglichen. Es folgt also  $LP(SUG \text{ Teilstudie 2.2}) > LP(SUG \text{ Teilstudie 2.1})$ .

Bei der Erstellung von Prompts ist ebenfalls die Spezifität der Prompts wichtig (Glogger u. a. 2009; Webb und Eves 2007). Hier ist die Frage, ob dies bei kollaborativer Reflexion in einer digitalen Forenumgebung (in dieser Teilstudie simuliert) ebenfalls ein relevanter Faktor ist. In dieser Teilstudie zielen alle drei Prompts darauf ab, dass Lösungsvorschläge eingebracht werden (siehe Tabelle 5-10). Jeder der Prompts hat einen leicht unterschiedlichen Fokus: Erstellung beliebiger Lösungsvorschläge (LP), die unabhängig von den existierenden Lösungsvorschlägen sein können, Auswahl eines Vorschlags und Argumentation für den Vorschlag (SLP) und Referenzieren eines Vorschlags (KLP). Wenn die Spezifität von Prompts eine Auswirkung hat, dann müsste in dieser Teilstudie ein Unterschied in den Antworten sichtbar werden.

*Hypothese 2: Prompts, die konkret nach Lösungen fragen (LP, SLP), generieren mehr Lösungsvorschläge als der Kontroll-Prompt (KLP).*

Diese Hypothese kann ebenfalls über den Code für allgemeine Lösungsvorschläge (SUG) überprüft werden. Die Operationalisierung dafür ist:  $LP(SUG) > KLP(SUG)$  und  $SLP(SUG) > KLP(SUG)$ .

Die zugrunde liegende Vermutung dieser Teilstudie ist, dass Prompts generell eine Wirkung haben. Dementsprechend sollten Prompts, die nach einem Bezug zu anderen Kommentaren in dem fiktiven Thema fragen, mehr Referenzen erzeugen als die übrigen Prompts.

*Hypothese 3: Prompts, die nach konkreten Referenzen zu existierenden Lösungen fragen (SLP), beinhalten mehr Referenzen zu diesen Lösungen als die Prompts für Lösungen (LP) bzw. der Kontroll-Prompt (KLP).*

Um diese Hypothese zu behalten, muss die Operationalisierung darstellen, dass mehr Beiträge mit Referenzen für den Prompt, der nach Bezug zu anderen Lösungen (SLP) fragt, vorhanden ist. Dazu müssen diese Bezüge explizit mit einem Code markiert werden. Dementsprechend wird die Hypothese wie folgt überprüft:  $SLP(Referenz\_existierende\_Lösungen) > LP(Referenz\_existierende\_Lösungen)$  und



SLP(Referenz\_existierende\_Lösungen) > KLP(Referenz\_existierende\_Lösungen). Diese Referenzen auf existierende Lösungen werden durch Codes in der qualitativen Inhaltsanalyse abgebildet (siehe Abschnitte 5.2.2.3 und 5.2.4).

#### 5.2.4 Deskriptive Analyse der Daten

Das Vorgehen bei der statistischen Analyse, sowie die Überprüfung der Voraussetzungen für die statistischen Verfahren ist in Anhang B beschrieben.

Vor der Analyse der Daten wurden die Antworten manuell und auf Basis der Gold-Standard-Frage (siehe Abschnitt 5.2.2.2) überprüft. Es wurden insgesamt 20 Antworten entfernt, in denen die Probanden nicht zum Thema passende Antworten gegeben haben. Eine weitere Antwort wurde entfernt, weil die Person laut Nennung des Alters minderjährig war. Beim Kontroll-Lösungs-Prompt wurden leicht mehr Antworten entfernt als bei den anderen beiden Prompts.

Erste Kalkulationen von Krippendorffs Alpha zeigten eine moderate Übereinstimmung zwischen den beiden Kodierenden (Hayes und Krippendorff 2007). Da sich diese Teilstudie überwiegend auf neue Lösungsvorschläge konzentriert, gibt es unter den Antworten wenige Codes für Fragen, Emotionen (EMO\_OWN oder EMO\_OTH), oder Interpretationen (INT), Lernen (S\_LOOP oder D\_LOOP) oder Wandel (CHANGE). Für die Codes für Erfahrung (EXP), Wissen (KNO), Ablehnung (DISAGR), Zustimmung (AGR), Hinweise (ADV), erfahrungsbasierte Lösungsvorschläge (SUG\_EXP), wissensbasierte Lösungsvorschläge (SUG\_KNO) beträgt der durchschnittliche Krippendorffs Alpha .85 (min. .76, max. .90). Die Codes EXP, KNO, ADV, SUG\_EXP und SUG\_KNO werden in dieser Analyse betrachtet. In dieser Teilstudie werden nur die Daten genutzt, bei denen die beiden Kodierenden übereingestimmt haben. Dies reduziert zwar leicht den Umfang des Datensatzes, erhöht aber die Datenqualität.

Der durchschnittliche Krippendorffs Alpha für die übrigen vier neuen Codes (siehe Abschnitt 5.2.2.3) beträgt .71 (min. .69, max. .74). Damit ist das  $\alpha$  leicht unter der Empfehlung von .800 (Krippendorff 2004). Der Krippendorffs Alpha für A\_DIFFERENT, der Code für den Lösungsvorschlag andere Aufgaben zu erledigen, beträgt lediglich .26. Da dieser unterhalb dem empfohlenen Wert von  $\alpha \geq .667$  liegt, wird dieser Code von den Analysen ausgeschlossen (Krippendorff 2004). Da die Werte teilweise nur 2x (A\_REQUESTER) oder 45x (A\_READ) vergeben wurden (im Vgl. zu A\_BREAK mit 245x Kodierungen), wurden die neuen Codes zu einer neuen Variablen (*ReferenceToOtherAnswers*) zusammengefasst. Ein  $\chi^2$ -Test zeigt keinen Zusammenhang zwischen den demografischen Variablen und dieser Variable. Ein weiterer  $\chi^2$ -Test zeigt

keinen Zusammenhang zwischen der neuen Variable und den reflexionsbezogenen Codes.

**Tabelle 5-12: Deskriptive Daten in Teilstudie 2.1**  
(alle Personen für Vergleich innerhalb der Teilstudie)

		<b>Kontroll- Lösungs-P.</b>	<b>Spezifischer- Lösungs-P.</b>	<b>Lösungs-P.</b>	<b>Alle</b>
<b>Allgemeines</b>	Anzahl Antworten	87	97	95	280
	Entfernte Antworten	13	3	5	20
<b>Zeit</b>	Durchschnitt	0:10:03	0:09:56	0:08:58	0:09:37
	Median	0:08:27	0:09:15	0:07:33	0:08:13
	St. Abw.	0:05:41	0:06:02	0:05:14	0:05:41
	Min.	0:02:12	0:01:18	0:02:11	0:01:18
	Max	0:29:33	0:29:50	0:25:43	0:29:50
<b>Geschlecht</b>	Männlich	49	54	46	150
	Weiblich	38	43	49	130
<b>Alter</b>	Durchschnitt	35,6	39,5	36,6	37,2
	Median	33,0	36,0	36,0	35,0
	St. Abw.	12,6	12,6	12,9	12,9
	Min.	19	21	21	19
	Max	72	71	63	72
	Keine Antwort	2	1	3	6
<b>Wörter</b>	Anzahl Wörter	6812	8037	7480	22397
	Durchschnitt	78,30	82,86	78,74	79,99
	Median	75,0	70,0	72,0	73,5
	St. Abw.	32,3	42,7	34,7	37,0
	Min.	26	19	21	19
	Max	197	218	191	218

Bei den Antwortzeiten fällt auf, dass bei dem Kontroll-Lösungs-Prompt die Personen am längsten für eine Antwort benötigen (siehe Tabelle 5-12). Die Antworten für den Lösungs-Prompt sind in der kürzesten Zeit verfasst und haben die geringste Standardabweichung. Es muss allerdings unterstrichen werden, dass diese Messungen nicht verlässlich sind, da es keinen Indikator gibt, inwieweit die Probanden während des Jobs abgelenkt waren, bspw. durch das Smartphone. Von daher muss diese Analyse vorsichtig interpretiert werden. Aufgrund der Streuung der Werte wurden diese in eine neue Variable gruppiert (analog wie in Teilstudie 2.1, siehe Abschnitt 5.1.4). Die Verteilung der Werte ist in Tabelle 5-13 dargestellt. Ein  $\chi^2$ -Test ergab, dass die neue Variable mit Zeit-Gruppen in keinem Zusammenhang mit den reflexionsbezogenen Codes EXP, KNO, ADV, SUG\_EXP und SUG\_KNO steht.

**Tabelle 5-13: Anzahl der Werte für jede Gruppe in der gruppierten Variable für die Zeit**

Gruppe	Anzahl Werte
0-240s	29
241-480s	106
481-720s	72
>720s	72

Ein  $\chi^2$ -Test zeigt, dass es einen signifikanten Zusammenhang zwischen der Anzahl der Wörter und den Codes für Erfahrungen (EXP), allgemeine Vorschläge (SUG) und erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) gibt (siehe Tabelle 5-14). In erster Linie lässt sich feststellen, dass ein Zusammenhang bei erfahrungsbasierten aber nicht bei wissensbasierten Codes vorhanden ist. Zusätzlich zeigt sich hier, dass wenn mehr Wörter genutzt werden, häufiger ein Code in den Texten vergeben wird. Der Anteil bei EXP und SUG sinkt bei zwei Stufen leicht. Letztlich zeigt sich, dass mehr Wörter keinen linearen Anstieg des Anteils nach sich zieht, wie oft ein Prompts einen der Codes nach sich zieht. Die Regressionsanalyse zeigt den Umfang der Rolle der Anzahl der Wörter in Abschnitt 5.2.5.

**Tabelle 5-14: Anteil, wie oft ein Prompt je nach Anzahl der Wörter einem Text hinzugefügt wurde**

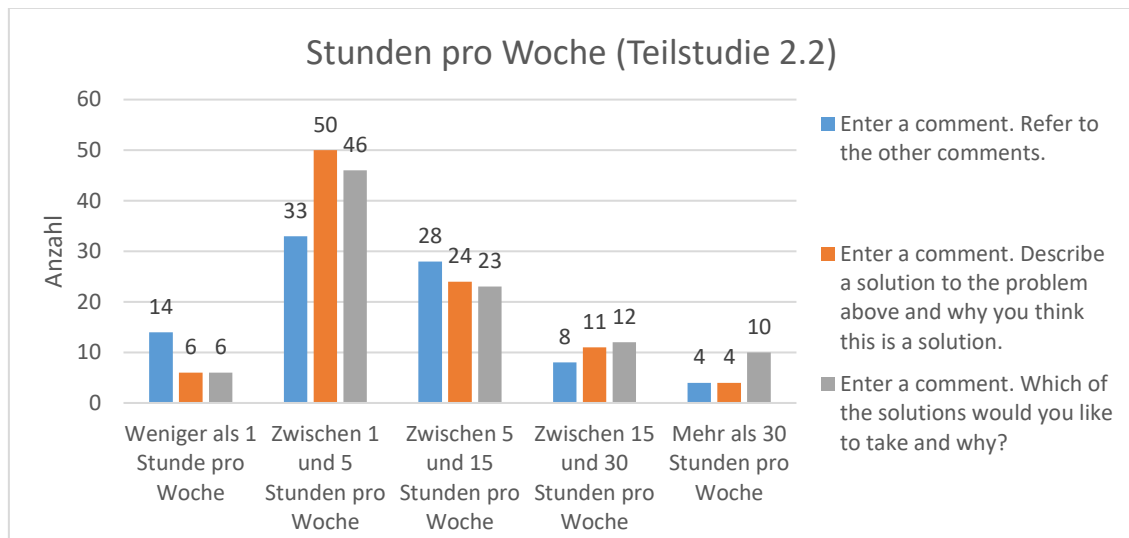
Code	0-40	41-80	81-120	121+	$\chi^2$ -Test
EXP	30,3 %	52,6 %	50,7 %	71,1 %	$\chi^2(3)=11,824, p=.008$
SUG	51,5 %	73,7 %	89,3 %	84,2 %	$\chi^2(3)=20,288, p<.001$
SUG_EXP	15,2 %	33,1 %	41,3 %	50 %	$\chi^2(3)=10,913, p=.012$

Bei der Verteilung der Geschlechter zeigt sich, dass in dieser Teilstudie die Probanden ebenfalls überwiegend männlich sind (53,5 % zu 46,5 %, siehe Tabelle 5-12). In anderen Studien findet sich ein überwiegender Anteil von Frauen (52 %) (Ross u. a. 2010). Hier haben alle Personen ihr Geschlecht angegeben. Ein  $\chi^2$ -Test zeigt keinen Zusammenhang zwischen dem Geschlecht und den reflexionsbezogenen Codes.

Das Alter ist relativ gleich verteilt, wobei die Probanden beim spezifischen Lösungs-Prompt leicht älter sind als die übrigen Probanden (siehe Tabelle 5-12). Insgesamt sechs Probanden haben ihr Alter nicht mitgeteilt. Für die Analyse wurde das Alter in Altersgruppen einsortiert, welche sich an Ross et al. orientieren (2010): 18-24, 25-34, 35-44, 45-54, 55-64, 65+. In einem  $\chi^2$ -Test wurde kein Zusammenhang zwischen den Altersgruppen und den reflexionsbezogenen Codes gezeigt.

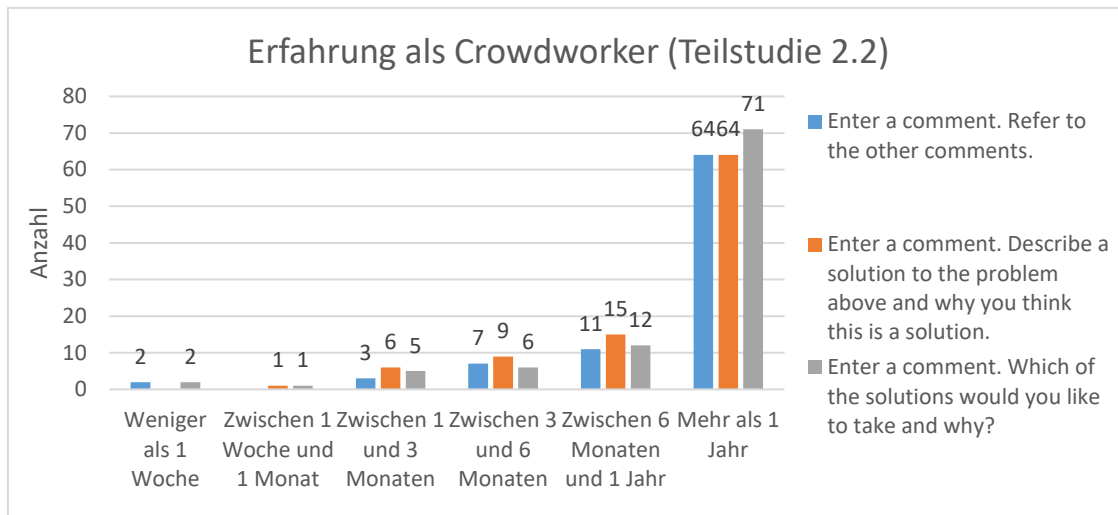
Abbildung 5-7 zeigt die wöchentlichen Arbeitsstunden der Probanden als Crowdworker. Die Zeiten sind bei allen Prompts ähnlich gleich verteilt, wobei in der Kategorie der Wochenarbeitszeit „zwischen ein und fünf Stunden“ weniger Probanden im Kontroll-Lösungs-Prompt sind. Der Großteil der Probanden arbeitet zwischen einer und 15

Stunden in Crowdsourcing. Laut eines  $\chi^2$ -Tests gibt es keinen Zusammenhang zwischen den reflexionsbezogenen Codes und der wöchentlichen Arbeitszeit als Crowdworker.



**Abbildung 5-7: Verteilung der wöchentlichen Arbeitsstunden der Crowdworker in Teilstudie 2.2 aufgeteilt nach Prompts**

Die meisten Probanden haben mehr als ein Jahr Crowdsourcing-Erfahrung (siehe Abbildung 5-8). Dies ist wenig verwunderlich, da analog zu Teilstudie 2.1 (siehe Abschnitt 5.1.4) nur Probanden mit Rang drei in die Teilstudie aufgenommen wurden, um die Datenqualität zu erhöhen (siehe Abschnitt 5.2.2.2). Da diese Stufe nicht sofort erreicht wird, hat der überwiegende Teil eine hohe Erfahrung als Crowdworker. Die Werte sind je Prompt-Kondition ähnlich verteilt. Bei einem  $\chi^2$ -Test werden nur Häufigkeiten  $\geq 5$  bewertet (Field 2009, 695–697). Da die Erfahrungsstufen „weniger als 1 Woche“ und „Zwischen 1 Woche und 1 Monat“ keine Werte haben, würden diese im  $\chi^2$ -Test nicht berücksichtigt werden. Für die Analyse wurde eine neue Variable (*ExpMoreThanAYear*) erstellt, die unterscheidet, ob der Proband mehr oder weniger als ein Jahr Erfahrung als Crowdworker hat. Es werden also fünf der sechs Kategorien zusammengefasst. Ein  $\chi^2$ -Test zeigt ein Zusammenhang zwischen der neuen Variable und dem Code für Erfahrung (EXP;  $\chi^2(1)=4,431$ ,  $p=,035$ ) und erfahrungsbasierten Lösungsvorschlägen (SUG\_EXP;  $\chi^2(1)=3,978$ ,  $p=,046$ ). Es wurden 77,8 % der Erfahrungen und 79,6 % der erfahrungsbasierten Lösungsvorschläge von Personen mit mehr als einem Jahr Erfahrung eingebracht. Da ca. 60% der Personen mehr als ein Jahr Erfahrung als Crowdworker haben, ist der Unterschied relativ gering. Dieser Zusammenhang kann dadurch erklärt werden, dass Personen, die weniger Erfahrung als Crowdworker haben, ein weniger umfangreiches Erfahrungswissen haben und dadurch weniger Erfahrungen beitragen können.



**Abbildung 5-8: Erfahrung der Crowdworke in Teilstudie 2.2 aufgeteilt nach Prompts**

In der Abbildung 5-9 ist der Bildungsgrad der Probanden je Prompt-Kondition dargestellt. Es zeigt sich, dass ähnlich wie in Teilstudie 2.1 (siehe Abschnitt 5.1.4) im Wesentlichen zwei Cluster vorhanden sind: Eine Gruppe hat einen High-School-Abschluss und (noch) keinen College-Abschluss und die andere Gruppe hat ein Bachelor- oder Masterabschluss. Zwischen den einzelnen Prompt-Konditionen sind die Daten im Wesentlichen ähnlich verteilt. Hier würden ebenfalls mehrere Werte in einem  $\chi^2$ -Test nicht berücksichtigt, da die Anzahl jeweils unter fünf liegt. Daher wurde eine neue Variable erstellt, die angibt, ob der Proband Akademiker ist, sprich einen Bachelor- oder Masterabschluss hat oder eine abgeschlossene Promotion hat. Der  $\chi^2$ -Test zeigt keine Zusammenhänge zwischen dem Bildungsgrad der Crowdworke und der reflexionsbezogenen Codes.

Abbildung 5-10 zeigt das Land der Crowdworke gemäß den Daten von Crowdfunder. Der überwiegende Teil der Crowdworke stammt aus Großbritannien und den USA. Dementsprechend spricht der überwiegende Teil der Probanden Englisch als Muttersprache. Weitere Länder, aus denen eine kleinere Anzahl Probanden stammen, sind Deutschland, Kroatien und Polen. Bei den meisten Ländern ist die Anzahl der Probanden zwischen den Prompt-Konditionen ähnlich verteilt. Ausnahmen bilden die USA und Großbritannien. Bei den USA sind im Vergleich zu den anderen beiden Prompts deutlich weniger Probanden in der Kontroll-Lösungs-Prompt-Kondition. Bei Großbritannien sind in dieser Kategorie deutlich mehr Nutzer. Zusammengefasst ist die Anzahl der Probanden, die Englisch als Muttersprache sprechen je Prompt-Kategorie allerdings ausgeglichen. Aufgrund der Verteilung der Werte wurde für die Analyse ebenfalls eine neue Variable angelegt (*NativeSpeaker*), die angibt, ob jemand Englisch als Muttersprache spricht (Gruppe 1: Großbritannien und USA; Gruppe 2: übrige Länder).

Ein  $\chi^2$ -Test zeigt, dass es einen Zusammenhang zwischen Erfahrungen (EXP;  $\chi^2(1)=5,626$ ,  $p=,018$ ) und erfahrungsbasierten Lösungsvorschlägen (SUG\_EXP;  $\chi^2(1)=22,095$ ;  $p<0,001$ ) gibt. Probanden, die Englisch als Muttersprache sprechen, haben 66,9 % der Erfahrungen und 78,8 % der erfahrungsbasierten Lösungsvorschläge eingebracht. In dieser Gruppe sind ebenfalls ca. 60 % der Werte. Dementsprechend ist hier der Unterschied nicht sonderlich groß. Eine Interpretation kann sein, dass es Muttersprachlern leichter fällt, eigene Erfahrungen schriftlich darzustellen. Dies kann mit der Anwendung des Kodierschemas zusammenhängen, bei dem ein Code nur vergeben wurde, wenn ein Wort existiert, auf das dieser Code exakt bezogen werden kann, um Interpretation zu vermeiden. Es ist möglich, dass es Nicht-Muttersprachlern schwieriger gefallen ist, die entsprechenden Wörter zu nutzen.

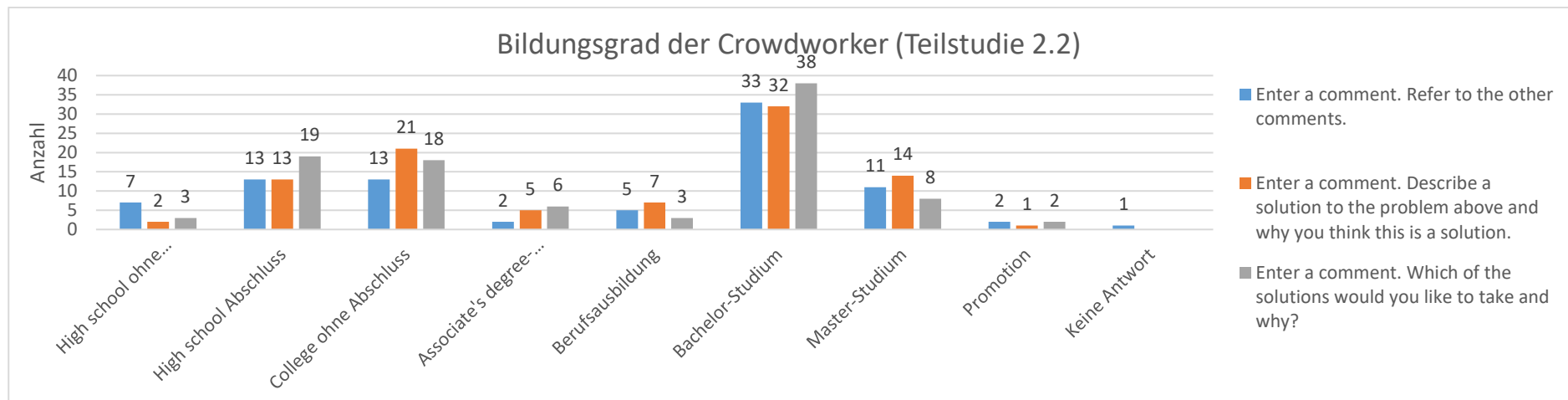


Abbildung 5-9: Bildungsgrad der Crowdworker in Teilstudie 2.2 je Prompt

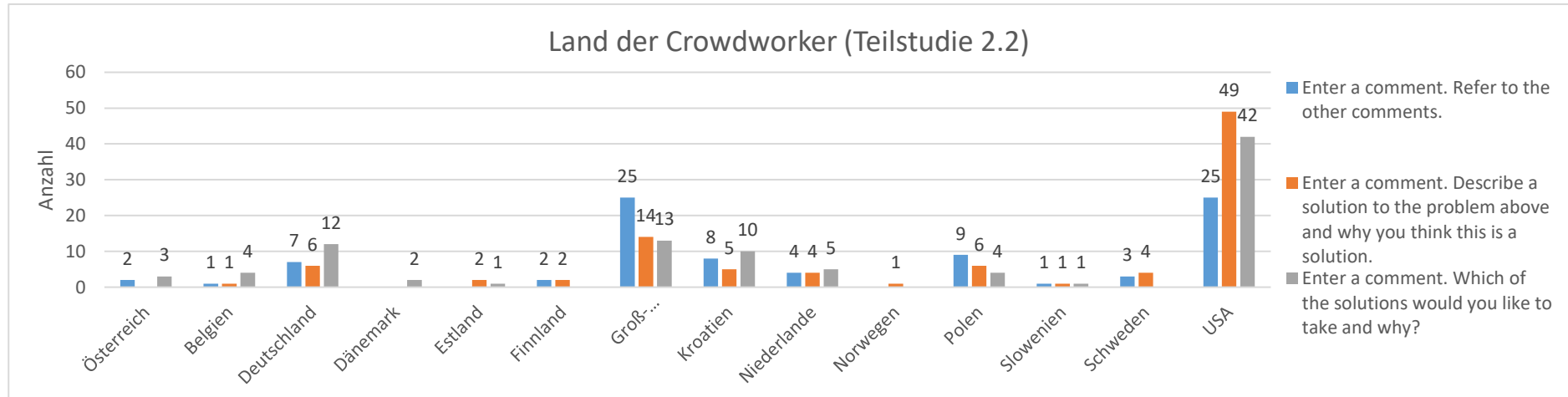


Abbildung 5-10: Land der Crowdworker in Teilstudie 2.2 laut Crowdfunder aufgeteilt nach Prompts

### 5.2.5 Qualitative Analyse

In mehreren logistischen Regressionen wird der Einfluss der Prompts und der vorherigen demografischen Daten auf die jeweiligen reflexionsbezogenen Codes überprüft.

Die erste Hypothese (siehe Abschnitt 5.2.3) vermutet, dass der Lösungs-Prompt zu mehr allgemeinen Lösungsvorschlägen (SUG) in einem Thema mit Diskussion (Daten aus dieser Teilstudie) als ohne Diskussion (Daten aus Teilstudie 2.1) führt. Zur Überprüfung von Hypothese 1 wurde zunächst ein  $\chi^2$ -Test durchgeführt. Dieser analysiert den Zusammenhang zwischen den Lösungs-Prompts und einerseits den Codes für Vorschlägen in dieser Teilstudie (mit Diskussion) und andererseits der vorherigen Teilstudie (ohne Diskussion). Der  $\chi^2$ -Test hat gezeigt, dass es weder für allgemeine Lösungsvorschläge (SUG), noch für erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) noch für wissensbasierte Lösungsvorschläge (SUG\_KNO) ein Zusammenhang existiert. Die Menge von 17 Probanden, die sowohl in Teilstudie 2.1 und Teilstudie 2.2 teilgenommen haben und in beiden Studien jeweils den Lösungs-Prompt gesehen haben, wurde nicht untersucht, da die Anzahl für eine sinnvolle statistische Analyse zu klein ist. Da kein Zusammenhang gefunden wurde, wird keine binär logistische Regression benötigt, um einen möglichen Effekt zu benennen.

In der zweiten Hypothese wird vermutet, dass Prompts, die versuchen Lösungen anzuregen (sowohl der Lösungs-Prompt (LP) als auch der Spezifische-Lösungs-Prompt (SLP)) mehr allgemeine Lösungen (SUG) hervorbringen als der Kontroll-Prompt (KP). Für die Analyse wird eine binär logistische Regression verwendet. Die abhängige Variable sind allgemeine Lösungsvorschläge (SUG) und die unabhängigen Variablen sind der Prompt, die demografischen Daten (Altersgruppe, Geschlecht, wöchentliche Arbeitsstunden, Academics, ExpMoreThanAYear, NativeSpeaker) sowie die Anzahl der Wörter. Das berechnete Modell ist signifikant für die Prädiktoren NativeSpeaker und die Anzahl der Wörter ( $\chi^2=35,153$ ,  $df=11$ ,  $p<,001$ ). Die übrigen Prädiktoren (Altersgruppe, Geschlecht, wöchentliche Arbeitsstunden, Academics, ExpMoreThanAYear und Prompt) fügen nicht signifikant zum Modell hinzu. Mit dem Modell werden 12,3 % (Cox & Snell  $R^2$ ) bzw. 18,6 % (Nagelkerke  $R^2$ ) der Varianz von allgemeinen Lösungsvorschlägen (SUG) erklärt. Dementsprechend erklärt das Modell nur einen eher kleinen Teil der Varianz der Lösungsvorschläge und es gibt weitere unbekannte Faktoren, die den Großteil der Varianz erklären. Im Modell werden 16,1 % der nicht-kodierten Texte (SUG) und 96,6 % der kodierten Texte von Lösungsvorschlägen korrekt klassifiziert. Insgesamt klassifiziert das Modell 78,1 % der Kodierungen korrekt. In der logistischen Regression wird versucht, ein Modell mit unterschiedlichen Prädiktoren zu finden, welche die



Kodierungen der abhängigen Variable korrekt klassifizieren (im Vgl. zu den beobachteten Kodierungen) (Vgl. Field 2009, 282–286). In diesem Regressionsmodell ist insbesondere die korrekte Vorhersage der kodierten Lösungsvorschläge zu einem sehr hohen Prozentsatz erfolgreich.

Tabelle 5-15 zeigt, dass die Wahrscheinlichkeit auf einen Lösungsvorschlag um den Faktor 2,4 steigt, wenn der Verfasser des Textes Englisch als Muttersprache spricht. Zusätzlich ist bei dieser Variable die Anzahl der Wörter wichtig und es lässt sich aus der Tabelle ableiten, dass die Wahrscheinlichkeit auf einen Lösungsvorschlag stark steigt, sobald mehr als 40 Wörter in dem Text vorhanden sind. Wenn ein Nutzer zwischen 81 und 120 Wörter für seinen Lösungsvorschlag aufwendet, ist die Wahrscheinlichkeit, dass ein Lösungsvorschlag vorhanden ist, im Vergleich zu 41-80 Wörtern, um den Faktor 3,2 höher. Hier scheint allerdings eine Grenze erreicht, da sehr viele Wörter (121+) keinen Effekt im Modell haben.

**Tabelle 5-15: Einflüsse der Variablen NativeSpeaker und Anzahl der Wörter auf den Code SUG**

Prompt	Vgl.	B (S.E.)	Wald	Sig.	EXP(B)	95 % CI für (EXP(B))	
						Lower	Upper
<b>Native Speaker</b>	Non-Native	,891 (.322)	7,654	,006	2,438	1,297	4,583
<b>W0-40</b>	W41-80	,923 (.457)	4,077	,043	2,517	1,027	6,167
<b>W0-40</b>	W81-120	2,090 (.589)	12,603	<,001	8,089	2,551	6,167
<b>W0-40</b>	W121+	1,648 (.656)	6,310	,012	5,198	1,436	18,108
<b>W41-80</b>	W81-120	1,167 (.448)	6,796	,009	3,213	1,336	7,729

Die dritte Hypothese vermutet, dass Prompts, die versuchen einen Bezug zu existierenden Lösungen anzuregen, zu Texten führen, in denen mehr dieser Referenzen auftauchen als bei den beiden anderen Prompts. Dementsprechend wird die kombinierte Variable als abhängige Variable geprüft, in der vermerkt ist, ob eine Referenz auf vorherige Antworten enthalten war. Die unabhängigen Variablen sind dieselben, wie in den vorherigen Kalkulationen. Das Modell ist nur für den Prädiktor Zeit, sprich die zur Erledigung des Jobs aufgewendete Zeit, signifikant. Wenn aber versucht wird, die einzelnen Gruppen der Zeit-Variable zu vergleichen ist das  $\chi^2$  des Modells nicht mehr signifikant. Dementsprechend wird an dieser Stelle der Schluss gezogen, dass es keinen Prädiktor für das Referenzieren existierender Lösungen gibt.

Zusätzlich wurden die reflexionsbezogenen Codes für Hinweise (ADV), erfahrungsbasierte (SUG\_EXP) und wissensbasierte Lösungsvorschläge (SUG\_KNO) untersucht. Für die beiden Codes ADV und SUG\_KNO gibt es kein Modell mit signifikanten Prädiktoren und auch ein  $\chi^2$ -Test zeigt keine Zusammenhänge. Für den

Code SUG\_EXP gibt es ein Modell, welches im Folgenden beschrieben wird. Bei erfahrungsbasierten Lösungsvorschlägen sind die Variablen NativeSpeaker und Anzahl der Wörter signifikante Prädiktoren im Modell ( $\chi^2=37,003$ ,  $df=11$ ,  $p<,001$ ). Die übrigen Variablen (Altersgruppe, Geschlecht, wöchentliche Arbeitsstunden, Academics, ExpMoreThanAYear und Prompt) sind keine Prädiktoren. Das Modell erklärt 12,9 % (Cox & Snell  $R^2$ ) bzw. 17,6 % (Nagelkerke  $R^2$ ) der Lösungsvorschläge. Im Modell werden 85 % der nicht-kodierten erfahrungsbasierten Lösungsvorschläge (SUG\_EXP) und 43,8 % der erfahrungsbasierten Lösungsvorschläge korrekt klassifiziert. Somit klassifiziert das Modell 70,3 % der Kodierungen korrekt. Tabelle 5-16 zeigt, dass die Wahrscheinlichkeit auf erfahrungsbasierte Lösungsvorschläge um Faktor 3,4 steigt, sobald der Verfasser Englisch als Muttersprache spricht. Die Anzahl der Wörter spielt ebenfalls eine Rolle und die Daten deuten darauf hin, dass die Wahrscheinlichkeit für erfahrungsbasierte Lösungsvorschläge um den Faktor zwei steigt, wenn mehr als 121+ Wörter geschrieben werden (im Vergleich zu 0-80 Wörter). Dementsprechend lässt sich ableiten, dass zumindest 80+ Wörter geschrieben werden sollten.

**Tabelle 5-16: Einflüsse der Variablen NativeSpeaker und Anzahl der Wörter auf SUG\_EXP**

Prompt	Vgl.	B (S.E.)	Wald	Sig.	EXP(B)	95 % CI für (EXP(B))	
						Lower	Upper
<b>Native Speaker</b>	Non-Native	1,224 (.313)	15,300	<,001	3,401	1,842	6,281
<b>W0-40</b>	W81-120	1,207 (.586)	4,236	,040	2,119	,718	6,255
<b>W0-40</b>	W121+	1,605 (.649)	6,122	,013	4,978	1,396	17,748
<b>W41-80</b>	W121+	,854 (.423)	4,076	,044	2,349	1,025	5,384

## 5.2.6 Diskussion

In diesem Abschnitt werden die Ergebnisse im Kontext der Hypothesen diskutiert und es werden die übrigen Ergebnisse interpretiert. Aus den Ergebnissen werden Hinweise für das Design von Tools zur Unterstützung von Reflexion abgeleitet.

### 5.2.6.1 Überprüfung der Hypothesen

Aus der qualitativen Analyse lässt sich entnehmen, dass die Prompts bei keinem reflexionsbezogenen Code ein signifikanter Prädiktor in den Regressionsmodellen waren. Somit können die beiden Hypothesen 1 und 2 nicht gehalten werden. Das Vorhandensein von existierenden Lösungen ist demnach kein Auslöser für mehr Lösungsvorschläge. Interessanterweise führen in diesem Setting die beiden Prompts für Lösungen (Lösungs-

Prompt und Spezifischer-Lösungs-Prompt) nicht zu mehr Lösungsvorschlägen als der Kontroll-Prompt. In dem Regressionsmodell, welches Einflüsse auf die Variable für Referenzen auf existierende Lösungen sucht (Hypothese 3), wurden keine signifikanten Prädiktoren gefunden. Das bedeutet, dass in diesem Setting die Formulierung der Prompts keinen Einfluss auf die Texte genommen hat.

Es werden alle drei Hypothesen verworfen. Hier lässt sich allerdings nicht zweifelsfrei beurteilen, ob das Ziel der Anregung von Lösungen mit den Prompts möglich ist (Z8). Die Ergebnisse der statistischen Kalkulation sagen lediglich aus, dass es keinen Unterschied zwischen den Gruppen gab, aber nicht, dass es keine Lösungen gab.

Überraschend ist, dass das Vorhandensein einer Diskussion in dieser Teilstudie im Vergleich zu der fehlenden Diskussion in der vorherigen Teilstudie nicht zu einem Unterschied bei der Menge der Lösungsvorschläge führt (Hypothese 3). Dieser fehlende Unterschied impliziert, dass die Anzahl der Lösungsvorschläge mit und ohne Diskussion ähnlich hoch ist. Dies passt nicht zur Literatur, in der häufig impliziert wird, dass zunächst eine Diskussion erfolgt, bevor Lösungsvorschläge generiert werden (Daudelin 1996; Høyrup 2004; Woerkom und Croon 2008; Raelin 2002; Dyke 2006; Hatton und Smith 1995). Eine daraus resultierende Hypothese ist, dass Unterstützungsfunktionen für Reflexion direkt von Beginn einer Diskussion an effektiv versuchen können Lösungsvorschläge anzuregen. Hier stellt sich allerdings die Frage, ob die Qualität der Lösungsvorschläge ohne zuvor stattfindende Diskussion gleichwertig ist.

In dieser Teilstudie haben die Prompts im Vergleich zur vorherigen Teilstudie leichte Formulierungsunterschiede aufgewiesen. Die Ergebnisse zeigen, dass diese geringen Unterschiede sich in diesem Studiendesign nicht auf die Reflexion bzw. die Texte auswirken. Dadurch das gezeigt wurde, dass im Vergleich zur vorherigen Teilstudie eine vergleichbare Menge Lösungsvorschläge verfasst wurde, bedeutet dies, dass in dieser Teilstudie die Prompts zu ähnlich waren, um einen messbaren Unterschied zu erzeugen. Daraus ergeben sich drei Interpretationen:

- a) Prompts können generell gut Konzepte anregen (Teilstudie 2.1), aber die Feinheiten in den Prompts schlagen sich nicht auf die geschriebenen Texte nieder (in dieser Teilstudie).
- b) Die Feinheiten der Prompts haben in einer laufenden Diskussion einen geringeren Einfluss.
- c) Die Probanden fanden die Konzepte in den Prompts (Auswählen einer existierenden Lösung oder Referenzieren einer Lösung) für die Konstellation in der Diskussion unpassend.

Renner et al. vermuten, dass Personen allgemein gerne Vorschläge zu einem Problem unterbreiten, ohne dass ein bestimmter Prompt zugegen sein muss (Renner u. a. 2019).

In dieser Teilstudie führen die beiden Lösungs-Prompts nicht zu mehr Lösungen als der Kontroll-Lösungs-Prompt. Das könnte vermuten lassen, dass die beiden Lösungs-Prompts eine geringere Wirkung hatten. Da jedoch die Texte in dem fiktiven Thema alle einen Lösungsvorschlag beinhalten, und der Kontroll-Lösungs-Prompt in dieser Teilstudie den Empfänger dazu aufgefordert hat sich auf die existierenden Antworten zu beziehen, kann es sein, dass dadurch implizit mehr Lösungen beim Kontroll-Lösungs-Prompt entstanden sind. Dieser Effekt könnte eine Erklärung für die fehlenden Effekte der Prompts sein. Eine andere Erklärung ist, dass Probanden leichter Lösungsvorschläge eingebracht haben, weil bereits andere Lösungsvorschläge in den Diskussionen existierten.

#### **5.2.6.2 Interpretation der übrigen Variablen**

Die Daten zeigen, dass die meisten demografischen Daten, wie Alter, Geschlecht, Bildungsgrad, wöchentliche Arbeitszeit oder die Erfahrung als Crowdworker in dieser Teilstudie keine Rolle spielen. Die einzige demografische Variable, die bei Lösungsvorschlägen bzw. bei konkret erfahrungsbasierten Lösungsvorschlägen eine Rolle spielte, ist die Angabe, ob die Probanden Englisch als Muttersprache haben. Offenbar fällt es diesen Personen leichter einen (erfahrungsbasierten) Lösungsvorschlag zu formulieren. Es sei angemerkt, dass die Daten von Crowdfunder genutzt werden, und daher dies Personen sind, die aus einem Land kommen, in dem Englisch die Muttersprache ist. Die Probanden können aus anderen Ländern zugezogene Personen mit anderer Muttersprache sein. Dennoch lässt sich ein Unterschied zwischen den Personen aus diesen Ländern und den übrigen Ländern feststellen.

Ein weiterer Faktor ist die Anzahl der Wörter, die bei Lösungsvorschlägen (SUG) und erfahrungsbasierten Lösungsvorschlägen (SUG\_EXP) eine Rolle gespielt hat. Hier zeigen die Daten auch, dass für einen Vorschlag ca. 80+ Wörter benötigt werden und das wenige Wörter (0-40) seltener Lösungsvorschläge in den Texten beinhalten. Aus diesen Werten lässt sich überlegen, ob es für Prompts sinnvoll sein kann, den Verfasser von Nachrichten darauf aufmerksam zu machen, falls der Beitrag eine niedrige Anzahl an Wörtern hat.

### **5.2.6.3 Möglichkeiten für das Design von Prompts**

Aus dieser Teilstudie lassen sich ein paar Überlegungen für das Design von Unterstützungsmöglichkeiten für kollaborative Reflexion ableiten.

Zum einen zeigt der fehlende Unterschied in der Menge der Lösungsvorschläge bei Themen mit und ohne Diskussion, dass Prompts bereits von Beginn an nach Lösungen fragen können, anstatt erst Erfahrungsaustausch anzuregen. Hier ist allerdings nicht analysiert, ob die Lösungen ohne vorherigen Erfahrungsaustausch, wie in der kollaborativen Reflexion üblich, qualitativ ähnlich zu den Lösungen nach vorherigem Erfahrungsaustausch sind. Dazu sind weitere Studien notwendig (siehe Abschnitt 5.3).

Aufgrund der Wichtigkeit der Anzahl der Wörter kann es sinnvoll sein, dass ein System entsprechende Prompts zu diesem Thema anzeigt. Beispielsweise kann ein Prompt den Verfasser eines Textes darauf aufmerksam zu machen, falls ein Beitrag sehr kurz ist und dass es sinnvoller sein kann, mehr Informationen hinzuzufügen, falls jemand einen (erfahrungsbasierten) Lösungsvorschlag kommunizieren möchte.

Zusätzlich hat diese Teilstudie unterstrichen, dass die Prompts dazu führen, dass die groben Konzepte, die in den Prompt-Texten genannt werden, beachtet werden (hier: Lösungsvorschläge). Die textuellen Feinheiten, dass man einen existierenden Lösungsvorschlag auswählen soll, wurden nicht beachtet. Hier ist beim Design von Unterstützungsmöglichkeiten für Reflexion wichtig, dass eine eindeutige Formulierung genannt wird. Unter Umständen können Designer dieser Unterstützungsmöglichkeiten nur auf einem größeren Level die Reflexion der Probanden lenken.

Der spezifische-Lösungs-Prompt und der Kontroll-Lösungs-Prompt fordern beide den Empfänger auf, eine der existierenden Lösungen auszuwählen bzw. zu referenzieren. Dies bedeutet, dass wenn diese beiden Prompts in der Praxis eingesetzt werden, dass die Software die Prompts erst dann liefern darf, wenn a) mehrere Antworten in einer Diskussion enthalten sind und b) es wahrscheinlich ist, dass in diesen Antworten Lösungsvorschläge enthalten sind.

### **5.2.7 Einschränkungen**

Die Einschränkungen aus Teilstudie 2.1 hinsichtlich der Crowdsourcing-Umgebung existieren auch in dieser Teilstudie (siehe Abschnitt 5.1.7).

Bei der Konfiguration der Teilstudie wurde eine Gold-Standard-Frage zur Qualitätssicherung implementiert. Diese stellt allerdings nur eine Frage zu dem fiktiven

Problem, was der Nutzer hatte, aber nicht zu den einzelnen Antworten, die in der Diskussion bereits enthalten waren. Dadurch lässt sich nicht zweifelsfrei garantieren, dass die Nutzer alle Antworten sorgfältig gelesen haben. Die Überprüfung der Variable für den Bezug auf existierende Antworten eignet sich dafür nur bedingt, da ein Crowdworker aus Versehen einen gleichen Vorschlag machen kann, ohne die vorherigen Antworten gelesen zu haben.

In diesem Aufbau waren alle Beiträge in der Diskussion bereits sehr lösungsorientiert (siehe Abbildung 5-6). Aus diesem Grund kann es den Probanden in dieser Teilstudie leichter gefallen sein, die eigenen Lösungsvorschläge einzubringen, und dies kann ein Grund sein, warum die Unterschiede zwischen den einzelnen Prompts nicht messbar waren. An dieser Stelle sind weitere Studien notwendig, um zu prüfen, wie die Prompts aus dieser Teilstudie in einer anderen Diskussionsumgebung wirken, bspw. bei mehreren Antworten, die unbegründete Hinweise (ADV), oder Wissen (KNO) oder Erfahrungen (EXP) beinhalten.

### **5.2.8 Zusammenfassung**

In dieser Teilstudie wurden die Effekte unterschiedlicher aber leicht ähnlicher Lösungs-Prompts auf die reflexionsbezogenen Codes untersucht. In dem Setting wurde diesmal ein Thema inklusive dreier Antworten in einer Diskussion simuliert. Dementsprechend unterscheidet sich das Setting für den Lösungs-Prompt in dieser Teilstudie von dem Setting des Lösungs-Prompts aus der vorherigen Teilstudie (siehe Abschnitt 5.1) nur dahingehend, dass in dieser Teilstudie Antworten in dem simulierten Thema enthalten waren. Im Wesentlichen liefert diese Teilstudie zwei interessante Erkenntnisse:

In dieser Teilstudie haben die Prompts keinen Einfluss auf die Lösungsvorschläge gehabt. Eine Interpretation ist, dass das simulierte Forenumgebung durch die vorhandenen Kommentare bereits zu stark durch Lösungsvorschläge geprägt war, als das Prompts, die auf Lösungsvorschläge abzielen, noch einen Unterschied machen. Dies impliziert, dass Prompts in der Praxis feingranulare Konzepte eventuell nur schlecht transportieren können oder auch, dass die Prompts sehr gut auf die existierende Situation in der Diskussion passen müssen. Beides sind Hypothesen, welche in Studien überprüft werden müssen.

Es wurden keine Unterschiede zwischen der Menge der Lösungsvorschläge bei dem Versuchsaufbau ohne Diskussionsbeiträge und mit Diskussionsbeiträgen gefunden. Daraus ergibt sich, dass es mit Prompts gleich gut möglich ist, von Anfang an in Online-Diskussionen nach Lösungsvorschlägen zu fragen. Hier ist allerdings nicht klar,

inwieweit es qualitative Unterschiede in den Lösungsvorschlägen gibt. Dies muss in einer weiteren Teilstudie geklärt werden (siehe Abschnitt 5.3).

### **5.3 Teilstudie 2.3: Die Qualität von Lösungsvorschlägen in kollaborativer Reflexion**

Ein Ergebnis aus der vorangegangenen Studie ist die Implikation, dass es keinen Unterschied gibt, ob ein Prompt direkt zu Beginn eines Themas oder erst in einer laufenden Diskussion nach Lösungen fragt. In den Ergebnissen der Teilstudie gab es keinen signifikanten Unterschied in der Häufigkeit der Codes für Lösungswissen. Offen ist jedoch, welchen Einfluss der Zeitpunkt der Prompts auf die Qualität der Lösungsvorschläge hat. In dieser Teilstudie sollen die Lösungsvorschläge der vorherigen beiden Studien systematisch bewertet werden, um die Qualität zu analysieren. Eine Klärung dieser Fragestellung hilft bei der spezifischeren Gestaltung von Unterstützungsmechanismen für kollaborative Reflexion.

#### **5.3.1 Ansatz**

Kollaborative Reflexion wird in der Literatur als Prozess in einer Gruppe beschrieben, der durch Erfahrungsaustausch, Diskussion, dem Hinterfragen von Annahmen, sowie dem Generieren, Diskutieren und Verwerfen von Ideen geprägt ist (Daudelin 1996; Høyrup 2004; Woerkom und Croon 2008; Raelin 2002; Dyke 2006; Hatton und Smith 1995). Es wird in der Literatur impliziert, dass die Teilnehmer kollaborativer Reflexion zunächst diskutieren und am Ende Lösungen entwickeln und deren Umsetzung planen. Dies liegt daran, dass während der Diskussion andere die Möglichkeit haben innerhalb der Reflexion Rückfragen zu stellen, um detaillierte Informationen zu dem Sachverhalt des Reflektierenden zu bekommen. Sobald Lösungsvorschläge direkt zu Beginn eingebracht werden, entfällt aber diese Diskussion. Dementsprechend liegt es nahe, dass den Theorien bzw. geplanten Maßnahmen dann *etwas* fehlen kann, um sie als produktives Ergebnis einer kollaborativen Reflexion zu sehen. Es wird vermutlich das Lernen aus sekundären Erfahrungen eingeschränkt (Dyke 2006). Die beiden vorangegangenen Studien haben gezeigt, dass es sowohl mit als auch ohne Diskussion bei einem Lösungs-Prompt vergleichbar viele Lösungsvorschläge gibt. Dieser Fund ist zunächst überraschend, da die Literatur impliziert, dass in der kollaborativen Reflexion vor der Lösungsfindung zunächst eine Diskussion stattfindet. Es stellt sich allerdings die Frage, inwieweit die Lösungsvorschläge, die durch einen Lösungs-Prompt angeregt werden, vergleichbar sind. Da Reflexion auf dem Austausch von Erfahrungen basiert, fokussiert sich diese Studie auf die Qualität von erfahrungsbasierten Lösungsvorschlägen.

Die Forschungsfrage in dieser Teilstudie lautet also: *Inwieweit wirkt sich das Vorhandensein von Diskussionsbeiträgen in einer kollaborativen Reflexion bei Nutzung eines Lösungs-Prompts auf die Qualität der Lösungsvorschläge aus?*

Viele Teilnehmer in Teilstudie 2.1 haben (siehe Abschnitt 5.1) auch beim Erfahrungs-Prompt (Z6) Lösungsvorschläge eingebracht. In dieser Teilstudie soll zusätzlich untersucht werden, ob es einen möglichen Qualitätsunterschied zwischen den Lösungen gibt, die durch den Lösungs-Prompt (Z8) und den Erfahrungs-Prompt (Z6) in Teilstudie 2.1 entstanden sind. Zur Analyse der Qualität werden unterschiedliche Kriterien erarbeitet anhand derer die Lösungsvorschläge bewertet werden. Die Beantwortung dieser Frage ermöglicht Rückschlüsse darauf, welche Prompts in einer Situation gezeigt werden können. Teilstudie 2.1 zeigt, dass in der Teilstudie der Lösungs-Prompt signifikant mehr Lösungsvorschläge hervorgerufen hat als die beiden anderen Prompts. Dies sagt aber nichts über die Qualität der Lösungsvorschläge aus. Falls bspw. der Erfahrungs-Prompt höherqualitative Lösungsvorschläge nach sich zieht, macht es ggf. mehr Sinn, diesen Prompt zu nutzen.

In dieser Teilstudie haben die Probanden einzelne Lösungsvorschläge aus den beiden vorherigen Studien bewertet. Das Ziel ist ein Vergleich der Qualität der Lösungsvorschläge anhand unterschiedlicher Kriterien (siehe folgenden Abschnitt 5.3.2.4), die entstanden sind, als es nur das Thema ohne Diskussionsbeiträge gab (Teilstudie 2.1, siehe Abschnitt 5.1) und den Lösungsvorschlägen, die durch Prompts in einer Umgebung mit drei Antworten innerhalb eines Themas (Teilstudie 2.2, siehe Abschnitt 5.2) entstanden sind. Es gab in beiden Studien sowohl unterschiedliche Prompts als auch einen Prompt, der gleichgeblieben ist (Lösungs-Prompt). Dadurch, dass der Lösungs-Prompt in beiden Studien identisch ist, ermöglicht dies einen Vergleich der Qualität der aus dem Prompt resultierenden Texte in einer Umgebung ohne Diskussionsbeiträge und in einer Umgebung mit Diskussionsbeiträgen. Dies soll in dieser Studie hauptsächlich untersucht werden.

Trotz der Fortschritte in der Forschung zu maschinellern Lernen lassen sich Ideen nicht ohne vorheriges manuelles Training evaluieren bzw. benötigen vorherigen Aufwand zur Vorbereitung (Christensen u. a. 2018). Daher wird diese Teilstudie ebenfalls in einer Crowdsourcing-Umgebung durchgeführt. Die Eigenschaften, sowie die Vor- und Nachteile so einer Umgebung sind im Abschnitt 3.5 näher beschrieben. In diesem Job bewerten die Probanden anhand der Kriterien die Lösungsvorschläge.

Da die Zielsetzung dieser Dissertation die Unterstützung von reflexivem Lernen ist, konzentriert sich diese Teilstudie auf erfahrungsbasierte Lösungen. Lösungen, die mit



allgemeinem Wissen begründet werden, zählen nicht zu Lernen aus Erfahrung und bleiben daher in dieser Teilstudie unberücksichtigt.

### **5.3.2 Methodik**

In diesem Abschnitt wird der konkrete Studienaufbau sowie die Herleitung der Kriterien zur Bewertung der Lösungsvorschläge dargestellt.

#### **5.3.2.1 Studiendesign**

Da sich das fiktive Szenario aus Teilstudie 2.1 (siehe Abschnitt 5.1.2) auf eine berufliche Situation eines Crowdworkers bezieht und die Probanden gebeten wurden an dieser Diskussion teilzunehmen, bietet es sich an, die erfahrungsbasierten Lösungsvorschläge von Crowdworkern bewerten zu lassen (Amabile 1982; Shah, Smith und Vargas-Hernandez 2003; Sukhov 2018). Kudrowitz und Wallace heben hervor, dass Laien Ideen oft nicht hinsichtlich ihrer Umsetzbarkeit bewerten können (Kudrowitz und Wallace 2013).

Zusätzlich zu den Bewertungen der Ideen und der Kontrollfrage werden demografische Daten erhoben. Darüber soll kontrolliert werden, ob die Gruppe der Probanden zwischen den Studien übereinstimmt oder ob es größere Abweichungen gibt.

In dieser Teilstudie sollen sich die Probanden auf die Bewertung von Lösungsvorschlägen konzentrieren und es wird als Kontext lediglich das fiktive Problem angezeigt. So werden auch die Lösungsvorschläge, die aus der Teilstudie 2.2 (Thema mit Diskussionsbeiträgen) resultieren, bewertet, da die Probanden während der Bewertung nicht durch die Lösungsvorschläge der Diskussionsbeiträge abgelenkt werden sollen. Die Beiträge, die unter dem Spezifischen-Lösungs-Prompt in Teilstudie 2.2 entstanden sind (siehe Tabelle 5-10 in Abschnitt 5.2.2.1), werden nicht berücksichtigt. Der Grund ist, dass viele dieser Beiträge eine abstrakte oder explizite Referenz enthalten, bspw. „*All of the commentaries above are helpful [...]*“ oder „*I prefer comment 3. [...]*“, und um diesen Lösungsvorschlag gut bewerten zu können, würde der Kontext der anderen Beiträge benötigt werden. Um in dieser Teilstudie die Umgebung für die Probanden einfach zu halten, wurde auf diese Komplexität verzichtet. Dementsprechend finden sich in dieser Teilstudie ausschließlich Lösungsvorschläge, in denen keine expliziten Referenzen auf die übrigen Diskussionsbeiträge enthalten sind.

### 5.3.2.2 Abbildung in FigureEight

In dem Job in FigureEight gibt es am Anfang eine kurze Erklärung der Ziele des Jobs (Bewertung von Lösungsvorschlägen) und des Hintergrunds (wissenschaftliche Studie). Der eigentliche Aufbau der Teilstudie ist in Abbildung 5-11 dargestellt.

#### Original Issue of a crowd worker:

Sometimes I get negative feedback on jobs I completed. Then the requester rejects my work and refuses payment although I did my best to answer the questions correctly and truthfully. This happened to me twice last week. Afterwards I get problems motivating myself because I invested so much time and effort. So I sometimes ask myself why should I take more jobs if I always get such negative experiences.

#### Suggestion in a reply by another crowdworker:

The only real solution is practise. I initially had similar experiences but over time I was better able to understand what the question was really expecting from me and my accuracy improved.

In the following, only rate the suggestion!

The suggestion is truly novel. (required)

strongly disagree

disagree

neither disagree nor agree

agree

strongly agree



[...]

If you had the issue, would you use the idea? (required)

- ☒ Yes
- ☐ Maybe
- ☐ No

Why would you use the suggestion? (Write at least two full sentences) (required)

Abbildung 5-11: Abbildung von Teilstudie 2.3 in FigureEight

Das fiktive Problem, das die Probanden in den beiden Studien bekommen haben, ist in der oberen Box („*Original issue of a crowd worker*“) dargestellt. Darunter befindet sich die zu bewertende Idee („*Suggestion in a reply by another crowdworker*“). Unterhalb dieses Textes befinden sich unterschiedliche Bewertungskriterien (siehe folgenden

Abschnitt 5.3.2.4). Unterhalb der Bewertungskriterien sind (in Abbildung 5-11 nicht sichtbar) dieselben demografischen Fragen, wie in den beiden vorherigen Teilstudien.

Die Probanden bewerten innerhalb des Jobs jeweils nur eine Idee, um zu vermeiden, dass die Bewertenden die unterschiedlichen Ideen miteinander vergleichen.

Die Limitierung der Laufzeit auf eine bestimmte Zeitspanne am Tag, wie es in den beiden vorherigen Studien genutzt wurde, war in bei FigureEight nicht möglich.

Obwohl manche Studien gezeigt haben, dass bereits mit sehr geringen Beträgen Ergebnisse erreicht werden können (Buhrmester, Kwang und Gosling 2011), ist die Bestrebung des Autors, eine faire Entlohnung sicherzustellen (Mason und Suri 2011; Martin u. a. 2014). Dazu wurde zunächst die Dauer des Jobs im Selbstversuch grob eingeschätzt, um einen Stundenlohn von zehn US-Dollar pro Stunde zu erreichen (Durchführung der Teilstudie Juni 2019). Es wurde zu Beginn ein kleiner Teil der Ideen bewertet, um sicherzustellen, dass die reale Bearbeitungszeit weiterhin im Rahmen bleibt. Pro Durchführung des Jobs wurden 0.40 \$ bezahlt.

Zusätzlich wurde bei der Erstellung des Jobs darauf geachtet, dass die Formulierungen für die Aufgaben möglichst konkret und kurz gehalten sind, sodass für die Crowdworker leicht ersichtlich ist, was die Aufgabe umfasst (Kittur, Chi und Suh 2008; Komarov, Reinecke und Gajos 2013; Martin u. a. 2014; Egelman, Chi und Dow 2014).

### **5.3.2.3 Auswahl und Vorbereitung der Daten**

Zur Bewertung der erfahrungsbasierten Lösungsvorschläge müssen die Daten aufbereitet werden. Im Versuch wird die Qualität der erfahrungsbasierten Lösungsvorschläge aus den vorherigen Studien (siehe Abschnitte 5.1 und 5.2) bewertet. Im Folgenden wird beschrieben, nach welchen Regeln, welche Daten ausgewählt bzw. bearbeitet wurden.

Da die Antworten in der vorangegangenen Teilstudie bereits inhaltlich kodiert wurden, ist bereits markiert, welche Beiträge einen erfahrungsbasierte, wissensbasierten oder allgemeinen Lösungsvorschlag enthalten und welche Beiträge über keine Lösungsvorschläge verfügen (siehe Abschnitt 5.2.2.3). Die in Teilstudie 2.1 (siehe Abschnitt 5.1) und Teilstudie 2.2 (siehe Abschnitt 5.2) entstandenen erfahrungsbasierten Lösungsvorschläge sollen dazu nach unterschiedlichen Kriterien bewertet werden. Es soll ein Vergleich zwischen den bewerteten erfahrungsbasierten Lösungsvorschlägen aus Teilstudie 2.1 (ohne vorherige Diskussion) und Teilstudie 2.1 (mit vorheriger Diskussion) durchgeführt werden. Aus der ersten Teilstudie (siehe Abschnitt 5.1) wurden alle

Antworten, die einen erfahrungsbasierten Lösungsvorschlag (SUG\_EXP) enthalten ausgewählt. Aus der zweiten Teilstudie (siehe Abschnitt 5.2) wurden alle Texte ausgewählt, die erfahrungsbasierte Lösungsvorschläge enthalten und bei einem der Prompts entstanden sind, bei denen nicht explizit ein anderer Kommentar in der Diskussion referenziert werden sollte.

Da in dieser Teilstudie nur die Qualität der erfahrungsbasierten Lösungsvorschläge bewertet werden soll, würde das Vorhandensein von Text zu anderen sich in der Diskussion befindlichen Antworten für die Bewertenden nur störend sein. Ebenfalls soll verglichen werden, wie sich die Qualität der Lösungsvorschläge, die bei einem Prompt für Lösungswissen (Lösungs-Prompt) entstanden sind, von der Qualität der Lösungsvorschläge unterscheidet, die bei einem der anderen Prompts entstanden sind.

Es wurden nur die Textstellen ausgewählt und bearbeitet, die von beiden Kodierenden mit dem entsprechenden Code kodiert wurden, um eine hohe Datenqualität sicherzustellen. Um erfahrungsbasierte Lösungsvorschläge zu bewerten, wurden nur Antworten ausgewählt, in denen der Code SUG\_EXP vorhanden war. Damit die Bewertung möglichst eindeutig ist und zusätzliche andere Vorschläge (SUG\_KNO) oder Hinweise (ADV) die Bewertenden nicht irritieren könnten, wurden diese aus den Antworten entfernt. Bei Hinweisen (ADV) ist dieses Entfernen oft einfach, weil es unbegründete Hinweise sind, die in vielen Fällen in der Form von „*don't give up*“ am Ende einer Antwort stehen. Das Entfernen hat keine Auswirkung auf den Zusammenhang bzw. die inhaltliche Aussage des Textes. Der Code SUG\_KNO wird immer in Verbindung mit dem Code KNO vergeben (siehe Abschnitt 3.4.1): Der mit KNO kodierte Teil markiert das allgemeine Wissen, was als Begründung angegeben wird und der Code SUG\_KNO markiert den Vorschlag an sich. Bei der Bearbeitung der Antworten wurde nach der folgenden Regel vorgegangen: Wenn der KNO-Teil nach Entfernen des SUG\_KNO-Teils noch inhaltlich im Kontext der üblichen Antwort Sinn ergibt, dann wurde er belassen. Falls nicht, wurde er ebenfalls entfernt. Das folgende Beispiel verdeutlicht dies:

*I definitely know what you mean, some of the tasks can have very frustrating results! Sometimes I will be doing something correctly, as per their rules, and after submitting it will tell me I'm wrong when I know that I'm not. Usually if it does that and kicks me off a task, I'll take a break and do something else, maybe try the task again at a later time to see if they've fixed their errors. **There are so many different Crowdflower jobs, just keep trying different ones and you'll end up finding a couple that are naturally a breeze to you***

Der fett gedruckt markierte Satzteil beschreibt das Wissen (KNO) und ist eine allgemeine Aussage, auf die der unterstrichene Lösungsvorschlag (SUG\_KNO) folgt. Nach Entfernen des Lösungsvorschlags steht der Satzteil allein, der besagt, dass Crowdfunder viele Jobs hat. Dieser passt jedoch nicht zu dem vorherigen Text, wo der Autor schreibt, dass er bestimmte Tasks in der Zukunft nach einer Pause fortsetzt. In sehr wenigen Fällen kam es vor, dass die Antwort durch das Entfernen des wissensbasierten Vorschlags (SUG\_KNO) ihren eigentlichen Charakter verloren hat, da der erfahrungsbasierte Vorschlag (SUG\_EXP) nur eine sehr knapp skizzierte Alternative am Ende der Antwort war. In diesen Fällen wurde die Antwort nicht zur Bewertung ausgewählt.

Es sei an dieser Stelle darauf hingewiesen, dass erfahrungsbasierte Lösungen nur als solche kodiert wurden, wenn im Text zweifelsfrei hervorgeht, dass es sich auf eine persönliche Erfahrung bezieht. Falls die persönliche Erfahrung im Text nur impliziert wird, dann wäre dies eine Interpretation des Textes, die fehlerhaft sein kann. Diese strengere Kodierung ermöglicht eine höhere Datenqualität (siehe Abschnitt 4.3.3).

Es gibt Antworten, in denen mehr als ein erfahrungsbasierter Lösungsvorschlag enthalten ist. Damit der Bewertende exakt weiß, welchen Lösungsvorschlag er bewerten soll, wurden diese Antworten in mehrere Antworten aufgeteilt. Andere Sätze, die Kontext gegeben haben (bspw. zustimmende Sätze am Anfang) wurden dann in allen aufgeteilten Antworten beibehalten. Wenn die Satzanfänge dadurch nicht nachvollziehbar waren bspw. „*Second, I suggest [...]*“ als Satzanfang für den zweiten Vorschlag), wurden diese umformuliert, ohne den inhaltlichen Sinn des Textes zu ändern.

Das folgende Beispiel zeigt eine Antwort, die einen Hinweis (ADV, fett gedruckt) und zwei erfahrungsbasierte Lösungsvorschläge beinhaltet (unterstrichen und doppelt unterstrichen):

*That is very frustrating and I would find it hard to get motivated if that was happening to me. **Have you tried contesting the rejection? I would outline why you feel the rejection of the job isn't fair while keeping an open mind about why the rejection happened.** I would also look at what type of jobs they are, if they are similar style jobs, I would maybe avoid that type for a while to keep myself from getting discouraged and work on jobs I am confident in my abilities on. You can also see if maybe you are doing too many jobs in row, I find that I need a break after a certain amount otherwise I become a little burnt out. Best of luck to you.*

In diesem Fall wird der Hinweis (fett gedruckt) entfernt. Da die beiden Lösungsvorschläge jeweils ein „also“ beinhalten, was auf jeweils andere Aussagen

hindeutet, werden diese entfernt. Der erste Satz bleibt enthalten, weil der Autor damit weiteren Kontext in die Antworten bringen wollte und der Satz ist dann in den beiden aus der Aufteilung resultierenden Antworten enthalten:

*That is very frustrating and I would find it hard to get motivated if that was happening to me. I would look at what type of jobs they are, if they are similar style jobs, I would maybe avoid that type for a while to keep myself from getting discouraged and work on jobs I am confident in my abilities on.*

*That is very frustrating and I would find it hard to get motivated if that was happening to me. You can also see if maybe you are doing too many jobs in row, I find that I need a break after a certain amount otherwise I become a little burnt out. Best of luck to you.*

Bei der Aufbereitung der Daten wurden vereinzelt kleinere Fehler in der eigenen Kodierung behoben, bspw. wenn ein eindeutiger Vorschlag nicht als solcher markiert war.

In Tabelle 5-17 ist eine Übersicht, welcher Datensatz aus den jeweiligen Studien, wie oft bearbeitet wurde und wie oft Antworten entfernt bzw. aufgeteilt wurden. Der Anteil an Beiträgen, die aufgeteilt werden müssen, ist im zweiten Versuch höher. Da sich dieser hauptsächlich mit Lösungswissen befasste, war dies zu erwarten. Der Anteil der entfernten Antworten ist als niedrig zu werten.

**Tabelle 5-17: Übersicht über die Bearbeitungsschritte in der Datenaufbereitung**

<b>Bearbeitungsschritt</b>	<b># Teilstudie 2.1</b>	<b># Teilstudie 2.2</b>
<b>Entfernt</b> (Kernbotschaft nicht mehr gegeben ohne Codes SUG_KNO oder ADV; oder keine Übereinstimmung auf Satzebene zwischen den Codern).	3	1
<b>Aufteilen:</b> Antworten mit mehreren SUG_EXP wurden auf mehrere Antworten aufgeteilt (ggf. mit Kontext).	3	13
<b>Problemlose Bearbeitung</b> (Keine Entfernung, kein Aufteilen)	23	23
<b>Daten insgesamt</b>	75	66

#### **5.3.2.4 Maße zur Bewertung von Lösungsvorschlägen**

Um eine Vergleichbarkeit der Lösungsvorschläge zu ermöglichen, werden in diesem Abschnitt Bewertungskriterien erarbeitet. Eine komplette Liste der Bewertungskriterien

findet sich am Ende des Abschnitts. Auf Basis dieser Bewertungskriterien werden die Lösungsvorschläge systematisch bewertet.

Aus der bestehenden Übersicht der Literatur (siehe Abschnitte 2.4 und 2.5) ist bekannt, dass kollaborative Reflexion den Austausch von Erfahrungen und das Diskutieren von Ideen beinhaltet (Dyke 2006; Nicholson u. a. 2003; Woerkom und Croon 2008; Kim und Lee 2002). Es ist wichtig zu beachten, dass die Zielsetzung von Reflexion die Erarbeitung von neuen Theorien über einen Sachverhalt oder eine neue Herangehensweise für eine Situation ist (Boud, Keogh und Walker 1985). Bei Boud et al. ist das Ergebnis der Reflexion mit einer klaren Handlungsabsicht für die Zukunft verbunden. Der Lösungsvorschlag innerhalb der kollaborativen Reflexion wird dann von jemandem aus der reflektierenden Gruppe eingebracht, um zu versuchen, das Problem eines anderen zu lösen. Derjenige, der einen Lösungsvorschlag einbringt, kann natürlich auch derjenige sein, für den das Problem besteht.

Ein Prototyp von Faltin et al. zeigt ein System mit dem Ideen, nach dem sie ausprobiert wurden, anhand von drei Dimensionen bewertet werden können: Zeit (bspw. Dauer in Minuten), Nutzen (bspw. Anzahl abgeschlossener Aufgaben), und allgemein der Grad, wie gut die Lösung funktioniert hat (Kriterien übersetzt aus Faltin, Schwantzer und Jung 2014). Da Ideen nach der Ausführung bewertet werden, passen diese Kategorien nicht für die geplante Teilstudie. Aus der Literatur zu Reflexion bzw. kollaborativer Reflexion sind keine weiteren spezifischen Kriterien bekannt, wie Lösungsvorschläge strukturiert sein sollten, um die reflektierende Person in ihrem Vorhaben zu unterstützen. Oft werden nur Situationen beschrieben, in der eine Gruppe reflektiert und Lösungen diskutiert (bspw. Glazer, Abbott und Harris 2004; Oliver, Omari und Herrington 1998).

Folgende Kategorien lassen sich theoretisch herleiten. Reflexion wird immer dann ausgelöst, wenn eine Situation anders verläuft, als derjenige es erwartet hat (Dewey 1933; Schön 1983; Kolb 1984; Boud, Keogh und Walker 1985). Dementsprechend hat der Reflektierende einen konkreten Problemkontext, zu dem er eine Lösung sucht. Eine dazu passende Lösung sollte generell *hilfreich sein*, um für den Reflektierenden von Nutzen zu sein. Das erste Kriterium ist *der Grad, wie hilfreich die Lösung ist*. Sobald andere in kollaborativer Reflexion Lösungsvorschläge einbringen, sollten diese auf den Problemkontext des Reflektierenden passen. Das zweite Kriterium ist demnach *der Grad, wie sehr die Lösung zum Problem passt*. Gleichzeitig sollte der Vorschlag so gestaltet sein, dass der andere ihn annimmt. Es genügt nicht, wenn der Lösungsvorschlag technisch das Problem lösen würde, aber er in der Realität selten akzeptiert werden würde. Dementsprechend soll ein Bewertungskriterium genutzt werden, das überprüft, *inwieweit der Bewertende den Lösungsvorschlag nutzen wollen würde* und *inwieweit der*

*Bewertende meint, dass der Lösungsvorschlag das eigene Verhalten beeinflusst, sollte er das Problem haben.* Ein Vorteil des Austausches in der kollaborativen Reflexion ist, dass man Vorschläge bekommen kann, an die man selbst nicht gedacht hat. Das wird in dem Kriterium abgebildet, *inwieweit der Vorschlag Aspekte enthält, an die man selbst nicht gedacht hat.* Dieses Kriterium grenzt sich von der Eigenschaft der Neuheit (s. u.) eines Vorschlags ab, der besagt, dass es wirklich eine echte neue Idee ist und nicht nur etwas ist, an das man selbst nicht aktiv gedacht hat. Um diesen Aspekte von rein ungewöhnlichen Vorschlägen, die eventuell nicht in ein existierendes Lösungsschema für diese Art Problem passen, zu anderen Aspekten zu trennen, soll zusätzlich ein Kriterium aufgenommen werden, das abbildet, *inwieweit der Vorschlag ungewöhnlich ist.*

Zusätzlich, wie oben erwähnt, ist die Zielsetzung bei Reflexion auf eine konkrete Handlungsabsicht bzw. eine Idee für eine Handlung hinzuarbeiten. Dementsprechend muss der Reflektierende am Ende der Reflexion idealerweise in der Lage sein, einen Lösungsvorschlag umzusetzen. Demnach ist das Kriterium, *der Grad, nach dem der Reflektierende auf Basis des Lösungsvorschlags handeln kann.* Dazu sollte die Lösung idealerweise ausreichend umfassend formuliert sein. Das dazugehörige Kriterium ist *der Grad, wie umfassend der Lösungsvorschlag formuliert ist.*

Vorschläge in kollaborativer Reflexion basieren häufig auf den eigenen Erfahrungen desjenigen, der den Vorschlag einbringt (Boerboom u. a. 2011). Da innerhalb der Teilstudie nur Lösungsvorschläge auf Erfahrungsbasis (Code SUG\_EXP; siehe Abschnitt 5.1.5 und Kodierschema in Abschnitt 3.4.1) und keine Lösungsvorschläge auf Wissensbasis (Code SUG\_KNO) untersucht werden, wird kein gesondertes Kriterium benötigt, das dies überprüft. Es kann angenommen werden, dass erfahrungsbasierte Lösungsvorschläge sehr glaubhaft sind, weil sie aus der persönlichen Erfahrung desjenigen stammen, der den Lösungsvorschlag macht. Ein Kriterium ist folglich *der Grad der Glaubwürdigkeit des Lösungsvorschlags.*

Die Literatur zu kollaborativer Reflexion ist hinsichtlich der Bewertung von Lösungsvorschlägen wenig aussagekräftig. Es lässt sich vermuten, dass eine Diskussion bzw. Erfahrungsaustausch im Rahmen einer Reflexion einem Lösungsvorschlag *etwas* hinzufügt, wodurch der Lösungsvorschlag besser bewertet sein müsste. Um die Ausgangsbasis zu erweitern, wurde an dieser Stelle ein Bereich herangezogen, in dem die Bewertung von Lösungsvorschlägen oder Ideen geläufiger ist: *Idea Management* und *Kreativität* (bspw. Dean u. a. 2006; Gerlach und Brem 2017; Kudrowitz und Wallace 2013). Häufig werden drei Phasen genannt: Ideensammlung, Auswertung und Auswahl von Ideen, bevor diese umgesetzt werden (Gerlach und Brem 2017). Da in der Teilstudie existierende Lösungsvorschläge hinsichtlich ihrer Qualität bewertet werden sollen,



fokussiert sich die Literatursauswahl auf die Studien, in denen eine Bewertung von Lösungsvorschlägen durchgeführt wurde. Die Literatur, die sich auf die Bewertung von Produkten bezieht, wird ausgeschlossen, da dort häufig auf Kriterien wie Nutzen für den Kunden und (technische) Umsetzbarkeit bzw. Überführbarkeit in ein kommerzielles Produkt genannt werden (Poetz und Schreier 2012; Xie und Zhang 2010). Diese Kriterien passen nicht zu dem fiktiven Szenario aus Teilstudie 2.1. Zusätzlich eignet sich Crowdsourcing sowohl für das Generieren als auch Evaluieren von Ideen (Wang, Wang und Tao 2017; Gerlach und Brem 2017; Kudrowitz und Wallace 2013).

Dean et al. haben eine umfassende Analyse von 90 existierenden Studien durchgeführt und beschreiben unterschiedliche Kategorien zur Bewertung von Ideen (vgl. im Folgenden Dean u. a. 2006). An dieser Stelle sind die übergeordneten Bewertungskategorien aufgelistet:

- „*Novelty*“: Der Grad der Neuheit der Idee und wie sehr die Idee existierende Paradigmen beeinflusst.
- „*Workability*“: Der Grad, wie einfach eine Idee in die Tat umgesetzt werden kann und der Grad, wie wenig sie gegen existierende Restriktionen verstößt (bspw. soziale, legale oder politische Rahmenbedingungen)
- „*Relevance*“: Der Grad, wie sehr die Idee zum genannten Problem passt und der Grad, wie effektiv sie das Problem lösen kann.
- „*Specificity*“: Der Grad, wie sehr eine Idee ausformuliert ist.

Die Kategorie *Novelty* kann für die geplante Teilstudie sinnvoll sein. Da kollaborative Reflexion auf dem Austausch von Erfahrungen basiert (siehe Abschnitt 2.5), liegt es nahe, dass die Ideen ebenfalls erfahrungsbasiert sind. Sobald der Lösungsvorschlag aus einer Erfahrung heraus begründet wird, hat derjenige diese Erfahrung gemacht und damit ist der Lösungsvorschlag nicht mehr als *neu* zu betrachten. Allerdings kann derjenige eine Idee haben, die auf eigenen Erfahrungen basiert, aber dennoch unterschiedlich von dem ist, was derjenige konkret erlebt hat und somit wirklich neu ist. Dementsprechend wird der Grad der Neuheit als Bewertungskategorie berücksichtigt.

Die übrigen Kategorien passen zu den zuvor aufgeführten theoretischen Überlegungen über gute Lösungsvorschläge im Rahmen kollaborativer Reflexion. Daher werden im Folgenden die Kategorien zur Umsetzbarkeit des Lösungsvorschlags, die Passung des Lösungsvorschlags zum Problem und die Verständlichkeit der Beschreibung des Lösungsvorschlags näher betrachtet.

Da das fiktive Szenario von dem Wunsch der persönlichen Verbesserung eines Crowdworkers handelt (siehe Abbildung 5-1 (2)) und nicht von der Erfindung eines neuen Produkts, ist ein Kriterium zum Grad, wie sehr der Lösungsvorschlag zu gegebenen Rahmenbedingungen passt, nicht relevant. Der Grad der Umsetzbarkeit des Lösungsvorschlags ist weiterhin interessant. Als Kriterium wird daher verwendet, *inwieweit der Empfänger sich zutraut, den Lösungsvorschlag in die Praxis umsetzen zu können*.

Die Kategorie der Relevanz umfasst zwei Dimensionen (vgl. im Folgenden Dean u. a. 2006), die sich beide eignen, um Lösungsvorschläge zu bewerten. Folglich werden als Kriterien sowohl der Grad, *wie sehr ein Lösungsvorschlag zum gegebenen Problem passt* und der Grad, *ob der Lösungsvorschlag das Problem lösen kann* ausgewählt. Ersteres wird bereits über die oben genannten Überlegungen abgedeckt. Das zweite Kriterium wird in die Liste der Kriterien aufgenommen.

Die Kategorie der *Specificity* gliedert sich bei Dean et al. in drei Unterkategorien auf, von denen eine später in der Evaluation der Bewertungskategorien verworfen wird (vgl. im Folgenden Dean u. a. 2006). Übrig bleiben die Dimensionen *Implicational explicitness* und *Completeness*. Die erste Bewertungskategorie beschreibt den Grad, wie sehr eine Verbindung zwischen der vorgeschlagenen Handlung im Lösungsvorschlag und dem erwarteten Ergebnis besteht. Dies prüft die interne Konsistenz des beschriebenen Lösungsvorschlags und wird ebenfalls verwendet. Die zweite Bewertungskategorie (*Completeness*) prüft, wie vollständig die Idee vor dem Hintergrund beschrieben ist, dass die Fragen, wer, was, wo, wann, warum und wie beantwortet sind. In einer Studie machen Dean et al. den Vorschlag, diese Kategorie auf einer Likert-Skala von 1-3 zu bewerten. Da die Bewertungskategorie insgesamt sechs Fragen an den Lösungsvorschlag stellt, erscheint eine einzelne kurze Likert-Skala hierfür unrealistisch. Das Kriterium passt bereits zu dem oben aus den theoretischen Überlegungen gewonnenen Kriterium, *wie konkret der Lösungsvorschlag formuliert ist*. Dadurch wird der Aspekt kollaborativer Reflexion abgedeckt, wonach Reflexion in eine Handlungsabsicht oder eine Handlung münden soll.

Um die allgemeine Qualität des Lösungsvorschlags zu messen, soll abgefragt werden, *inwieweit der Bewertende den Lösungsvorschlag weiterempfehlen würde bzw. inwieweit der Bewertende der Meinung ist, dass der Lösungsvorschlag sein eigenes Verhalten beeinflussen könnte, sollte er das Problem haben*.

Als weitere zusätzliche, objektive Kriterien lässt sich analysieren, *wie viele Wörter in einer Idee benutzt wurden und wie lange jemand an einer Idee gearbeitet hat* (Ye u. a.

2016). Beide Kriterien benötigen keine Bewertung in diesem Crowdsourcing Job, da die Daten bereits in den vorherigen Teilstudien erfasst wurden. Es ist zu vermuten, dass beide Kriterien zusammenhängen, da jemand, der eine lange Beschreibung einfügt, länger daran arbeitet.

### 5.3.2.5 Anwendung der Bewertungskriterien

Bei der Bewertung von vielen Lösungsvorschlägen ist es wichtig, die Anzahl der Bewertungskriterien pro Lösungsvorschlag im Rahmen zu halten, da ansonsten eine Bewertung sehr lange dauern kann (Kudrowitz und Wallace 2013). Bei einer niedrigen Anzahl zu bewertender Lösungsvorschläge können komplexe Bewertungskriterien genutzt werden (Kudrowitz und Wallace 2013). Der Wunsch dahinter ist, dass eine feingranulare Bewertungsmöglichkeit eine höhere Qualität der Bewertung ermöglicht (Blohm u. a. 2009). Blohm et al. impliziert, dass es bei vielen umfangreichen Kriterien (in der Publikation: 15 Kriterien je 1-7 Likert-Skala) hilfreicher sein kann auf binäre Skalen (bspw. „sofort realisierbar“ und „realisierbar in zwei Jahren“ (Blohm u. a. 2009, 12)) zu wechseln. Daher wird im Folgenden versucht einen sowohl kleineren Umfang an Kriterien bzw. eine kleinere Spanne der Likert-Skala zu wählen, um eine Balance zwischen Detailgrad in der Datenerhebung und der guten Bewertbarkeit durch die Teilnehmer zu ermöglichen.

Bei der Bewertung einzelner Kategorien gibt es in der Literatur keine einheitlichen Standards. In Tabelle 5-18 ist ein Überblick über den Umfang der Skalen bei der Bewertung einzelner Kategorien:

**Tabelle 5-18: Spanne von Bewertungskriterien in der Literatur**

<b>Umfang Likert-Skala</b>	<b>Studie(n)</b>
1-3 bzw. 1-4	(Dean u. a. 2006)
1-5	(Blohm u. a. 2009 Abbildung 2)
1-7	(Ardaiz-Villanueva u. a. 2011; Blohm u. a. 2009, 7)
1-10	(Christiaans 2002; Sukhov 2018)
1-100	(Magnusson, Netz und Wästlund 2014)

Eine Studie hat ergeben, dass die durchschnittlichen Bewertungen aus 5er und 7er Likert-Skalen vergleichbar sind und nur minimal höher als eine 10er Skala liegen (Dawes 2008). Gleichzeitig ist das Problem sehr großer Skalen (wie bspw. die 100er Skala bei Magnusson, Netz und Wästlund 2014), dass der Unterschied zwischen einzelnen Werten unpräzise wird (Krosnick und Fabrigar 2012). Auf Basis einer Literaturübersicht fassen Krosnick und Fabrigar zusammen, dass eine 5er oder 7er Skala in Bezug auf Validität und Reliabilität im Vergleich zu kürzeren oder längeren Skalen besser ist. Sie empfehlen

weiterhin verbale Beschriftungen anstatt Zahlenwerte zu verwenden. Daraus folgend wird im weiteren Verlauf eine 5er Likert-Skala mit verbalen Beschriftungen für die Bewertungskriterien der Lösungsvorschläge genommen.

Um eine möglichst objektive Bewertung der Kriterien sicherzustellen, sollen die Ideen von mehreren Personen bewertet werden. Die Anzahl der Personen, die eine Idee bewerten, ist in der Literatur stark unterschiedlich. Bei Ideenbewertungen in einem kommerziellen Kontext reicht die Anzahl von vier bis elf Personen (Magnusson 2009; Blohm u. a. 2010; Ye u. a. 2016). Im akademischen Kontext wurden zwischen zwei bis 96 Personen zur Bewertung von Ideen eingesetzt (Dean u. a. 2006; Kudrowitz und Wallace 2013; Christiaans 2002; Sukhov 2018). Es ist festzustellen, dass die beiden Studien mit sehr vielen Personen (34 bei Christiaans (2002) und 96 bei Sukhov (2018)) mit Studenten bzw. Bürgern über Umfragen gearbeitet haben. Kudrowitz und Wallace halten fest, dass zwei Personen zu wenig sind (Kudrowitz und Wallace 2013). Als Wert für die Teilstudie werden fünf Personen je Lösungsvorschlag geplant, da diese Zahl im Rahmen der meisten Studien liegt. Sollte sich die Bewertungen je Lösungsvorschlag zu stark unterscheiden, werden bei Bedarf weitere Bewertungen durchgeführt, um einen einheitlichen Wert zu erzielen. Da es sich um eine sehr subjektive Bewertung handelt, werden alle Bewertungen miteinbezogen. Die Bewertung der Probanden ist bspw. im Vergleich zur inhaltlichen Kodierung der Texte aus den vorherigen Studien weniger objektiv, weniger strikt und es fehlen bspw. Beispiele, anhand derer die Probanden die Bewertung vornehmen sollen. Aus den Gründen wird keine Übereinstimmung der Bewertenden auf Basis von bspw. Krippendorffs-Alpha genutzt (Hayes und Krippendorff 2007). Für die Analyse wird eine numerische Abbildung der Likert-Skalen genutzt.

Somit ergeben sich die folgenden Bewertungskriterien, die mit einer 5er Likert-Skala (strongly disagree, disagree, neither disagree nor agree, agree und strongly agree) abgefragt werden:

- K1. The suggestion is truly novel.
- K2. The suggestion is unusual.
- K3. The suggestion contains elements I did not think of myself.
- K4. The scope of the suggestion fits the issue.
- K5. It is likely that the suggestion solves the issue.
- K6. The suggestion is helpful.
- K7. The suggestion and its reasoning are believable.
- K8. The suggestion is clearly stated (e.g., it says what to do when).

K9. I'm confident that I could put the suggestion into practice.

K10. If I had the issue, the suggestion most likely would have an impact on my behavior.

K11. I would recommend the solution to others.

Für das Bewertungskriterium, ob der Bewertende den Lösungsvorschlag nutzen würde, wenn er das Problem hätte, wird eine 3er Likert-Skala verwendet (yes, no, maybe). Diese wird mit einem Textfeld kombiniert, um eine Begründung des Bewertenden abzufragen. Auf diese Weise können Kontextinformationen zum Lösungsvorschlag mit aufgenommen werden, um eine qualitative Bewertung zu ermöglichen:

K12. If I had the issue, I would use the suggestion.

Zusätzlich gibt es ein weiteres Bewertungskriterium, nach dem die Lösungsvorschläge evaluiert werden: die Anzahl der Wörter.

K13. Die Anzahl der Wörter im Lösungsvorschlag.

Ein Vergleich der Zeit ist nicht sinnvoll, da die Probanden im Versuchsaufbau von Teilstudie 2.2 (siehe Abbildung 5-6) zusätzlich drei neue Diskussionsbeiträge lesen mussten, welche in Teilstudie 2.1 noch nicht vorhanden waren. Demzufolge ist zu vermuten, dass die Probanden in dem Versuch mehr Zeit für den Lösungsvorschlag benötigten als im Versuch ohne die Diskussionsbeiträge.

### **5.3.2.6 Qualitätssicherung in der Crowdsourcing-Umgebung**

Die Crowdsourcing-Plattform bietet bei Erstellung eines Jobs mehrere Ränge an Expertise der Crowdworker zur Auswahl an, aus denen gewählt werden kann. Ein höherer Rang bedeutet, dass der entsprechende Crowdworker mehr Erfahrung hat. Ein niedrigerer Rang als Voraussetzung für den Job bietet den Vorteil, dass mehr Personen diesen Job bearbeiten können und er daher schneller bearbeitet wird. Andererseits wird mit einem höheren Rang an Erfahrung sichergestellt, dass die Ergebnisse eine höhere Qualität haben (Egelman, Chi und Dow 2014). Für diesen Job wurde der höchste Rang an Erfahrung der Crowdworker ausgewählt<sup>25</sup>.

Es wurde wie in den beiden vorherigen Teilstudien zur Sicherstellung des Sprachverständnisses eine Beschränkung der Herkunftsländer genutzt. Diese Teilstudie wurde auf die folgenden Länder beschränkt: Schweden, Niederlande, Singapur,

---

<sup>25</sup> <https://success.figure-eight.com/hc/en-us/articles/203219195-Guide-To-Contributors-Channels-Page>  
(Letzter Zugriff am 12.08.2019)

Norwegen, Dänemark, Südafrika, Luxemburg, Finnland, Slowenien, Deutschland, Belgien und Österreich. Diese Länder haben eine Bewertung von „*Very High Proficiency*“ im English Language Proficiency Index<sup>26</sup>. Zusätzlich wurden noch die Länder ausgewählt, in denen English die Muttersprache bzw. Amtssprache ist: USA, Großbritannien, Kanada und Australien. Da in vergangenen Studien asiatische Crowdworker im Vergleich schlechtere Arbeitsergebnisse erzielt haben, wurde Singapur ausgeschlossen (Kazai, Kamps und Milic-Frayling 2012). Neben der Sprache und dem Rang an Erfahrung in der Crowdsourcing-Plattform gab es keine weiteren Kriterien für die Auswahl der Probanden.

Die Nutzung von Gold-Standard-Fragen (siehe Abschnitt 3.6.7) ist bei typischen Umfrage-Jobs oder Jobs mit vielen multiple-choice Antworten, wie bei der Bewertung von Ideen anhand mehrerer Kriterien, schwierig (Gadiraju u. a. 2015). Die Herausforderung besteht darin, dass der Autor der Studie vorher nicht wissen kann, wie anwendbar eine der Ideen aus Sicht der Crowdworker ist. Daher wurde bei der Erstellung von Gold-Standard-Fragen darauf geachtet, dass nur sehr eindeutige Ideen in das Set von Gold-Standard-Fragen aufgenommen werden, bspw. wo die Beschreibung keinerlei Hinweise auf die Herangehensweise zur Umsetzung schließen lässt. Eine automatische Generierung von Gold-Standard Fragen ist in diesem Kontext nicht möglich.

Eine Gold-Standard-Frage zu stellen ist in diesem Job schwierig, da der Schwerpunkt hauptsächlich auf der Bewertung der Ideen liegt und diese in jedem Job-Durchlauf unterschiedlich sind. Die Anzahl der Wiederholungen des Jobs ist auf fünf pro Person begrenzt, um potenziellen Schaden von Crowdworkern mit Täuschungsabsicht zu minimieren, aber um den Job dennoch in annehmbarer Zeit durchzuführen. Da die Teilstudie über mehrere Tage lief und keine neuen Bewertungen hinzukamen, wurde nachträglich die Anzahl auf acht Bewertungen pro Person erhöht. Da die Instruktionen nur knapp beschreiben, dass man den gezeigten Lösungsvorschlag bewerten soll, welcher einerseits klar gekennzeichnet ist und sich jede der Bewertungskategorie eindeutig auf den Lösungsvorschlag bezieht, wurde auf die Erstellung von Instructional Manipulation Checks verzichtet (siehe Abschnitt 3.6.9). Dementsprechend müssen die Daten manuell auf Nutzer mit Täuschungsabsicht überprüft werden.

Gleichzeitig wurde in der Crowdsourcing-Plattform eine Mindestdauer für die Bearbeitung gesetzt. Wird diese unterschritten, wird die Antwort automatisch verworfen.

---

<sup>26</sup> <https://www.ef.com/wwen/epi/> (Letzter Zugriff am 12.08.2019)

Diese Maßnahme dient dazu die sogenannten *fast-deceiver* auszuschließen (siehe Abschnitt 3.6.7). Die Mindestdauer für eine Durchführung des Jobs beträgt 45 Sekunden.

### 5.3.2.7 Testlauf

Es wurde ein Test mit zehn Dokumenten à fünf Bewertungen (siehe Abschnitt 5.3.2.4) durchgeführt, um einerseits zu überprüfen, ob das Fehlen einer Gold-Standard-Frage problematisch ist, da Umfrage-Jobs an sich schwierig zu kontrollieren sind (siehe Abschnitt 3.5). Andererseits sollte geprüft werden, ob die eingeschätzte Zeitdauer bzw. die daraus resultierende Bezahlung passend ist.

Bei FigureEight ist eine Bewertung der Jobs möglich. Von den 50 durchgeführten Bewertungen haben Nutzer in fünf Fällen den Job bewertet. Es zeigt sich, dass der Job insgesamt gut zu verstehen ist und einfach durchzuführen ist (siehe Tabelle 5-19). Es wurde bewertet, wie fair die Testfragen sind. Da keine vorhanden sind, ist dieser Wert nicht relevant. Die Zeit der Antworten liegt durchschnittlich bei 3:24 Minuten (Mdn. = 2:54 Min.; SD = 2:20 Min.). Die Bezahlung ist auf ca. 2,5 Minuten Arbeit ausgelegt (0,40 \$ Bezahlung mit Ziel von zehn US-Dollar pro Stunde). Da die durchschnittliche Zeit dort sehr dicht an der avisierten Zeit liegt und die Bezahlung als gut bewertet wird, wird die Höhe der Entlohnung beibehalten.

**Tabelle 5-19: Bewertung des Probelaufs in Teilstudie 2.3 (n=5)**

Insgesamt	Klarheit der Instruktionen	Einfachheit des Jobs	Bezahlung
5/5	4,6/5	4,6/5	4,8/5

Eine einfache Durchsicht der Antwortverteilungen pro Nutzer je Fragekategorie ergab, dass die Nutzer unterschiedliche Bewertungen vergeben und keine Täuschungsabsicht vorzuliegen scheint. Diese Bewertung ist schwierig vorzunehmen, da es sich um subjektive Bewertungen von Ideen handelt und der Datensatz mit fünf Antworten pro Nutzer relativ klein ist. Es gibt zwei Bewertungen, in denen die Nutzer in allen Kategorien den gleichen Wert markiert haben. Diese Bewertungen stammen jedoch von zwei unterschiedlichen Nutzern, die mehrfach teilgenommen haben und in anderen Bewertungen differenzierte Antworten gegeben haben. Zusätzlich zeigt der Freitextkommentar bei einem der Nutzer, dass der Bewertende die Lösung nicht mochte. Die Zwischenergebnisse sind positiv und die Teilstudie bedarf keiner Veränderung vor der eigentlichen Durchführung.

Die Daten aus dem Testlauf werden später mit den übrigen Daten zusammengefasst, da es hinsichtlich des Studienaufbaus keinen Grund gibt, diese zu verwerfen. Die

teilnehmenden Crowdworker können im weiteren Studienverlauf nur teilnehmen, wenn sie bisher weniger als acht Einheiten bewertet haben (siehe Abschnitt 5.3.2.6).

### **5.3.3 Bereinigung der Studiendaten vor der Analyse**

Bevor die Analyse begonnen wurde, wurden die Daten auf Qualität untersucht. Nach dem ersten Durchlauf der Teilstudie wurde festgestellt, dass es vier Arten von Problemen mit den Daten gab:

Zunächst ist aufgefallen, dass es drei Personen möglich war, denselben Lösungsvorschlag zweimal zu bewerten. Dieses Problem wurde gelöst, indem die doppelte Bewertung aus dem Datensatz gelöscht wurde und die Analyse auf vier Bewertungen für den Lösungsvorschlag durchgeführt wurde.

Das zweite Problem sind Personen gewesen, die von Bewertung zu Bewertung unterschiedliche Daten angegeben haben. Ein Grund kann sein, dass sich die Personen verklickt haben, oder dass sich mehrere Personen einen Account geteilt haben. Mehrere Personen hatten bspw. acht Bewertungen abgegeben und bei einer davon war bspw. bei der Erfahrung ein anderer Wert angegeben als bei den anderen sieben Bewertungen. Die Daten dieser Personen wurden beibehalten. Die Möglichkeit acht, statt fünf Bewertungen abzugeben wurde erst nach einer Laufzeit der Teilstudie von ca. einer Woche konfiguriert. Mehrere Personen haben erst fünf Bewertungen abgegeben und dann zu dem späteren Zeitpunkt weitere drei Bewertungen. Bei diesen Bewertungen gab es bspw. einen Unterschied in der Erfahrung. Eine Person hat zunächst drei bis sechs Monate Erfahrung angegeben und bei der späteren Durchführung sechs Monate bis ein Jahr angegeben. Diese Entwicklung ist plausibel. Es hat sich ebenfalls häufiger die wöchentliche Arbeitszeit in der Angabe der demografischen Daten unterschieden. Daraus lässt sich zunächst ableiten, dass die Probanden diese Frage sehr auf die aktuelle Woche beziehen und nicht als eine allgemeine Frage verstehen. Die Daten dieser Personen wurden beibehalten. Bei allen Personen wurde zusätzlich geprüft, ob die Antwort auf die Frage, warum sie den Lösungsvorschlag nutzen wollen, sinnvoll auf die Bewertung und den Text passt. Wenn dies der Fall war, wurden die Daten beibehalten. Es gibt insgesamt vier Personen, die bei mehreren demografischen Daten unterschiedliche Werte bei unterschiedlichen Bewertungen angegeben haben. Eine Person hat bspw. bei fünf Bewertung fünf unterschiedliche Altersangaben getätigt. Bei diesen vier Personen ist zu vermuten, dass diese den Job nicht ernsthaft durchgeführt haben. Diese Daten wurden verworfen.



Das dritte Problem ist eine Person, die über insgesamt elf Benutzerkonten an dem Job teilgenommen hat. Die demografischen Daten waren über alle Bewertungen gleich, die ID der Benutzerkonten war an mehreren Stellen bei der achtstelligen Benutzerkennung sequenziell. Der letzte Beleg ist der Fakt, dass die exakt gleiche textuelle Antwort bei der Bewertung eines Lösungsvorschlags von mehreren Konten gegeben wurde. Da diese Person mehrere Lösungsvorschläge teilweise vier Mal bewertet hat und die unterschiedlichen Benutzerkonten widersprüchliche Bewertungen der Kriterien vorgenommen haben, wurden alle Daten dieser Benutzerkonten verworfen. Die betreffenden Benutzerkonten wurden gesperrt und der Plattform gemeldet.

Das vierte Problem waren Probanden, die den Job sehr schnell abgeschlossen haben. Für den Versuch wurden ca. 2,5 Minuten veranschlagt (=150 Sekunden). Es gab insgesamt sechs Probanden, die durchschnittlich für eine Bewertung unter 100 Sekunden benötigt haben. Dadurch lässt sich implizieren, dass diese Probanden die Aufgabe nicht gelesen haben könnten. Drei dieser Probanden haben widersprüchliche demografische Angaben getätigt und wurden daher entfernt. Die anderen drei Personen haben plausible Antworten gegeben und die textuelle Antwort auf die Frage, ob die Probanden die Idee nutzen wollen war plausibel. Die Daten dieser drei Personen wurden beibehalten.

Nachdem mehrere Personen und einige Bewertungen aus dem Versuch entfernt wurden, wurde eine Liste aller Lösungsvorschläge erstellt, die weniger als vier Bewertungen haben. Für diese Lösungsvorschläge wurden zwei Runden neuer Jobs erstellt und es wurden neue Daten gesammelt. Durch die neuen Runden gibt es acht Crowdworker, die zwischen 9 und 16 Lösungsvorschläge bewertet haben. Da die schriftlichen Bewertungen vertrauenswürdig waren und die demografischen Daten über mehrere Tage, an denen diese Personen an der Teilstudie teilgenommen haben, gleich blieben, wurden diese Personen zugelassen. Insgesamt haben 134 unterschiedliche Crowdworker an der Teilstudie teilgenommen. Es wurden 22 Bewertungen ausgeschlossen und 753 behalten. Es wurden je fünf Bewertungen für 155 Lösungsvorschläge im Job in Auftrag gegeben.

### **5.3.4 Demografische Daten der Probanden**

Nach der Bereinigung der Daten (siehe Abschnitt 5.3.3) blieben 753 Bewertungen von 134 Crowdworkern übrig. Ähnlich wie bei dem Probelauf waren die Probanden mit dem Job zufrieden, wie Tabelle 5-20 zeigt. Es haben lediglich 30 der 134 Teilnehmer an der Teilstudie eine Bewertung des Jobs abgegeben. Aus dieser Tabelle und der Tatsache, dass 112 Personen über fünf Mal und häufiger an dem Job teilgenommen haben, zeigt sich, dass die Bezahlung des Jobs fair ist und der Job ansprechend gestaltet war.

**Tabelle 5-20: Abschließende Beurteilung des Jobs durch die Crowdworker**  
(n=30 von 134 Personen im Job)

Insgesamt	Klarheit der Instruktionen	Einfachheit des Jobs	Bezahlung
4,8/5	4,8/5	4,6/5	4,6/5

Bevor die Daten ausgewertet werden, erfolgt zunächst eine Betrachtung der demografischen Daten der Probanden. Aus Tabelle 5-21 ergibt sich, dass insgesamt 20 Personen bereits an den früheren Crowdsourcing Studien teilgenommen haben, wobei eine Person davon an allen drei Studien teilgenommen hat. Eine Überprüfung der Daten hat ergeben, dass niemand seinen eigenen Lösungsvorschlag bewertet hat. Es wurden alle Einträge bereinigt, in denen eine Person eine Idee mehr als einmal bewertet hat.

#### 5.3.4.1 Demografische Analyse der Bewertenden

Bei manchen Berechnungen der  $\chi^2$ -Tests zeigte sich, dass eine Personengruppe häufig einen Extremwert wählt (*strongly disagree* bzw. *strongly agree*), während die andere Personengruppe häufig die anliegenden Werte (*disagree* bzw. *agree*) wählt. Dadurch kamen in der 5er-Likert-Skala häufig statistisch signifikante Zusammenhänge zustande, die inhaltlich nicht sinnvoll erklärbar sind. Zusätzlich haben mehrere Kategorien in einem der äußeren Werte eine niedrige Häufigkeit, was Auswirkung auf die statistischen Verfahren haben kann (bspw. 17 Werte bei *strongly disagree* K7 oder 26 Werte bei *strongly agree* in K3; jeweils von 753 Werten). Eine Möglichkeit damit umzugehen ist das Zusammenfassen der 5er-Likert-Skala zu einer 3er-Likert-Skala (bspw. Rink, Tricker und Harvey 2007). Dabei wurde die Mitte unverändert beibehalten und die jeweiligen Außenbereiche zu neuen Werten (negativ (*strongly disagree* und *disagree*) und positiv (*strongly agree* und *agree*)) zusammengefasst wurden, was in anderen Analysen zu ähnlichen Ergebnissen führt (Jae Jeong 2016). Auf diese Weise lässt sich das zuvor beschriebene Phänomen in der Analyse umgehen. Alle Instanzen, in denen die 3er-Likert-Skala genutzt wurde, sind im Folgenden explizit markiert.

Die Probanden in dieser Teilstudie sind durchschnittlich knapp 41 Jahre alt (Mdn.=39,5; SD=14,1) (siehe Tabelle 5-21). Damit ist das Alter der Probanden grob vergleichbar mit den vorherigen Teilstudien. In allen Teilstudien zeigt sich eine moderate Verteilung des Alters anhand der Standardabweichung. Eine Person hat sowohl in Teilstudie 2.1 als auch in Teilstudie 2.2 teilgenommen, wodurch insgesamt 20 Personen an einer früheren Teilstudie teilgenommen haben. Aufgrund der niedrigen Anzahl der Teilnehmer wurde keine gezielte Analyse der Teilgruppe bzw. kein Vergleich durchgeführt.

**Tabelle 5-21: Demografische Informationen über die Probanden in Teilstudie 2.3 (n=134)**

<b>Kategorie</b>		<b>Anzahl</b>
<b>Allgemeines</b>	Entfernte Antworten	22
<b>Geschlecht</b>	Männlich	60
	Weiblich	73
	Keine Antwort	1
<b>Alter</b>	Durchschnitt	40,8
	Median	39,5
	St. Abw.	14,1
	Min.	19
	Max	84
	Keine Antwort	2
<b>Teilnahme an vorherigen Experimenten</b>	Teilstudie 2.1	12
	Teilstudie 2.2	9
	Insgesamt	20

Ein  $\chi^2$ -Test zeigt mehrere signifikante Verteilungen der Bewertungen innerhalb der Altersgruppen (siehe Tabelle 5-22). Aus der Tabelle lässt sich entnehmen, dass die Personen in den Altersgruppen 17-24 und 55-64 am ehesten geneigt sind, einen Lösungsvorschlag so zu bewerten, dass er einen Einfluss auf das eigene Verhalten nimmt (K10). Personen im Alter zwischen 25 und 54 bewerten dies wesentlich seltener. Bei den jüngeren Personen kann das als mangelnde Erfahrung interpretiert werden. Dies passt zu dem hohen Wert für die Bewertung von neuen Inhalten bei den Lösungsvorschlägen in der jungen Altersgruppe. Ältere Personen haben Ideen eher als klar verständlich markiert (K8), als die Personen zwischen 17-34. Bei der Kategorie, ob eine Idee leicht in die Praxis überführbar ist (K9), haben ältere Personen eher zugestimmt als jüngere Personen (mit Ausnahme der Kategorie 65+, aber in dieser Kategorie sind nur wenige Personen enthalten; ein Zusammenfassen der Altersgruppe 55-64 und 65+ zeigt den gleichen signifikanten Zusammenhang). Es finden sich zwar viele Unterschiede, die statistisch signifikant sind, die sich aber in der Praxis nicht erklären lassen. Daher ist diese Tabelle nur mit Vorsicht zu interpretieren.

**Tabelle 5-22: Verteilung der Bewertungen innerhalb der Kategorien  
gruppiert nach Altersgruppe (3er-Likert-Skala)**

Rating	Alter	$\chi^2$ (df=10)	negative	Neither nor	positive
<b>Impact on my behaviour (K10)</b>	17-24	$\chi^2=19,520$ , p=,034	26,2 %	16,4 %	57,4 %
	25-34		34,4 %	23,9 %	41,8 %
	35-44		29 %	25,8 %	45,2 %
	45-54		31,4 %	23,6 %	45 %
	55-64		20,3 %	16,4 %	63,3 %
	65+		27,6 %	20,7 %	51,7 %
<b>Put into practice (K9)</b>	17-24	$\chi^2=21,304$ , p=,019	14,8 %	14,8 %	70,5 %
	25-34		20,9 %	10,4 %	68,7 %
	35-44		15,1 %	21 %	64 %
	45-54		10 %	15 %	75 %
	55-64		8,6 %	14,1 %	77,3 %
	65+		20,7 %	17,2 %	62,1 %
<b>Elements I didn't think of (K3)</b>	17-24	$\chi^2=40,341$ , p<,001	57,4 %	26,2 %	16,4 %
	25-34		68,2 %	12,9 %	18,9 %
	35-44		60,2 %	20,4 %	19,4 %
	45-54		77,9 %	7,1 %	15 %
	55-64		74,2 %	6,3 %	19,5 %
	65+		41,4 %	17,2 %	41,4 %
<b>Clearly stated (K8)</b>	17-24	$\chi^2=25,963$ p=,004	18 %	19,7 %	62,3 %
	25-34		22,4 %	14,9 %	62,7 %
	35-44		11,3 %	11,3 %	77,4 %
	45-54		11,4 %	17,1 %	71,4 %
	55-64		7 %	12,5 %	80,5 %
	65+		10,3 %	17,2 %	72,4 %
<b>Novel (K1)</b>	17-24	$\chi^2=23,754$ , p=,008	31,1 %	14,8 %	54,1 %
	25-34		44,8 %	18,4 %	36,8 %
	35-44		44,1 %	26,9 %	29 %
	45-54		47,9 %	25 %	27,1 %
	55-64		47,7 %	26,6 %	25,8 %
	65+		34,5 %	24,1 %	41,4 %

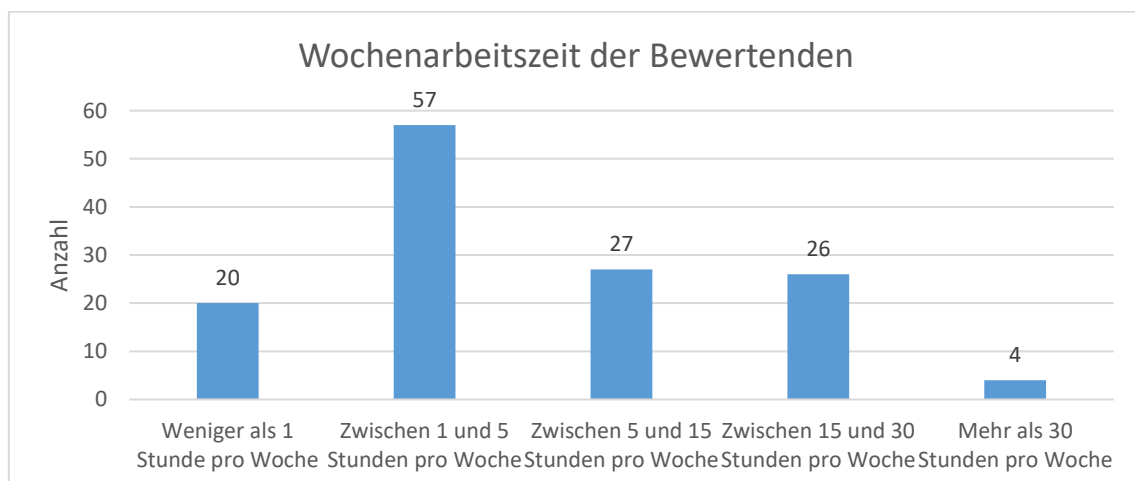
In dieser Teilstudie sind 45 % der Probanden männlich und 51 % der Probanden weiblich, wobei eine Person kein Geschlecht genannt hat. Dieser Wert ist vergleichbar mit anderen Studien, in denen der Anteil von weiblichen Probanden 52 % entsprach (Ross u. a. 2010). Tabelle 5-23 zeigt in einem  $\chi^2$ -Test einen signifikanten Zusammenhang zwischen den Variablen Geschlecht und der Bewertung für ungewöhnliche Ideen (K2). Die Daten implizieren, dass Männer öfter als Frauen einen Lösungsvorschlag als nicht ungewöhnlich bewerten und gleichzeitig, dass Frauen eher als Männer einen Lösungsvorschlag als ungewöhnlich bewerten. Der Zusammenhang ist nur sehr knapp

signifikant und sollte daher nicht überinterpretiert werden. In der 3er-Likert-Skala findet sich dieser Zusammenhang nicht.

**Tabelle 5-23: Verteilung der Bewertung ungewöhnlicher Ideen für Männer und Frauen  
(5er-Likert-Skala)**

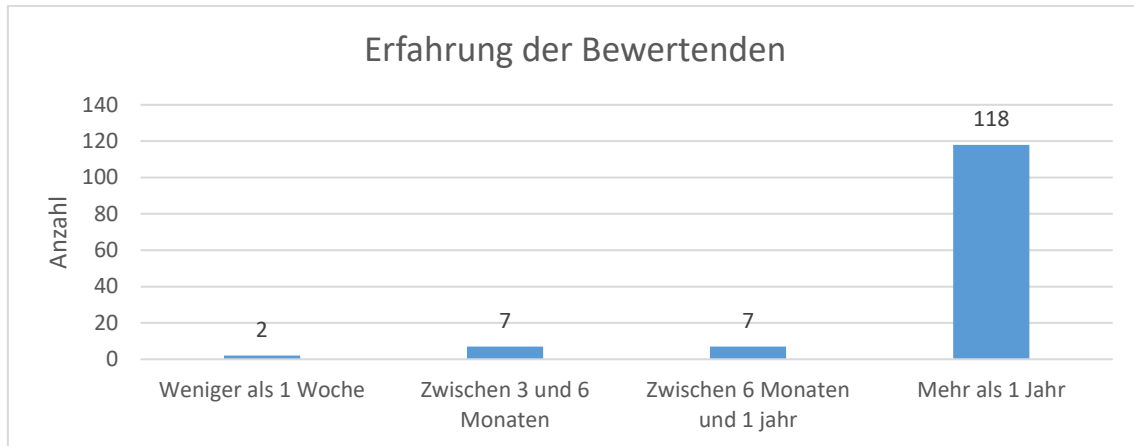
Rating	Geschlecht	$\chi^2$	Strongly disagree	disagree	Neither nor	agree	Strongly agree
<b>Unusual (K2)</b>	Weiblich	$\chi^2(4)=9,533$ , $p=,049$	20 %	45,9 %	15,4 %	15,4 %	3,3 %
	Männlich		28,2 %	41,8 %	16,7 %	11,5 %	1,9 %

Abbildung 5-12 zeigt, dass der überwiegende Teil der Probanden zwischen einer und fünf Stunden pro Woche als Crowdworcker arbeitet. Die Antworten zur wöchentlichen Arbeitszeit von Probanden, die an unterschiedlichen Zeitpunkten an dieser Teilstudie teilgenommen haben, waren unterschiedlich. Dies impliziert, dass die Probanden diese Frage sehr auf die aktuelle Woche bezogen haben (siehe Abschnitt 5.3.3). Diese Abbildung zeigt also eine Momentaufnahme und wird daher nicht in den demografischen Daten berücksichtigt.



**Abbildung 5-12: Wöchentliche Arbeitsstunden der Probanden in Teilstudie 2.3 (n=134)**

Ähnlich wie in den vorherigen Studien haben die Probanden in dieser Teilstudie überwiegend mehr als ein Jahr Erfahrung (Abbildung 5-13). Dies war zu erwarten, da für die Teilnahme an der Teilstudie nur Personen der höchsten Erfahrungsstufe zugelassen waren. Dementsprechend ist niemand in der Kategorie zwischen einer Woche und drei Monaten vertreten. Aufgrund der niedrigen Zahlen für die Personen mit weniger als einem Jahr Erfahrung, werden diese analog zu den beiden vorherigen Studien zu einer neuen Variable zusammengefasst (*ExpMoreThanAYear*): Diese drückt aus, ob jemand mehr oder weniger als ein Jahr Erfahrung hat.



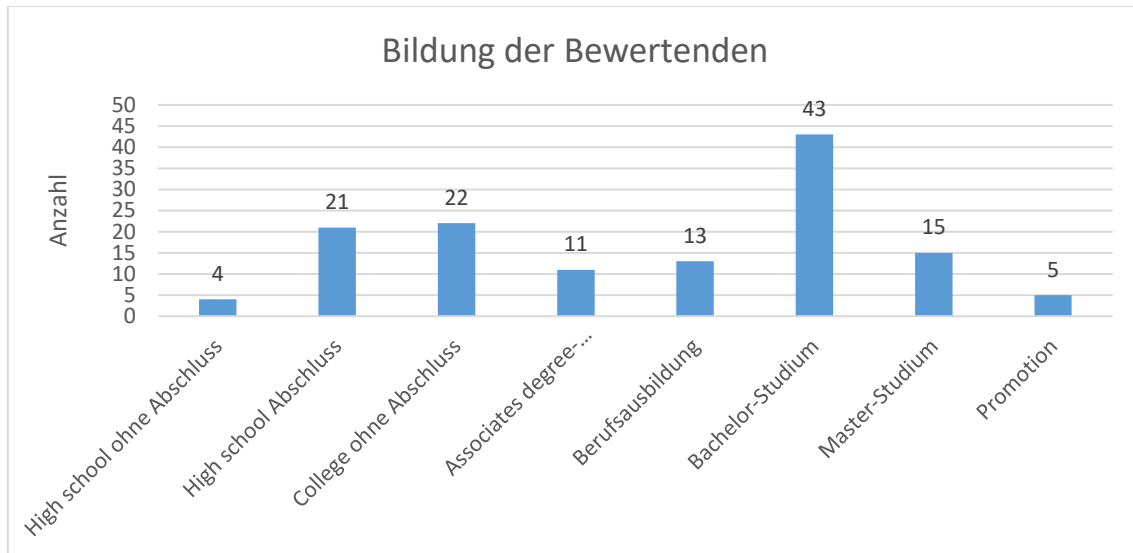
**Abbildung 5-13: Erfahrung der Probanden mit Crowdsourcing in Teilstudie 2.3 (n=134)**

Ein  $\chi^2$ -Test weist mehrere Zusammenhänge zwischen der Erfahrung und den Bewertungskategorien auf (siehe Tabelle 5-24). Es zeigt sich, dass Personen mit mehr Erfahrung dem Lösungsvorschlag kritischer gegenüberstehen, was den Grad betrifft, wie sehr die Lösung das Problem löst (K5). Nachvollziehbar ist, dass Personen mit weniger Erfahrung häufiger zustimmen, dass Elemente enthalten sind, an die sie selbst nicht gedacht haben (K3). Wer mehr Erfahrung hat, bewertet Ideen häufiger als klar verständlich (K8). Es kann sein, dass diese Personen das Problem kennen, zu dem Lösungsvorschläge entwickelt werden sollten, und daher den Lösungsansatz leichter verstehen. Zusätzlich bewerten weniger erfahrene Personen die Lösungsvorschläge eher als neu (K1). Dies ist ebenfalls aufgrund der fehlenden Erfahrung nachvollziehbar. Die Bewertung bei der Kategorie für ungewöhnliche Antworten (K2) ist nicht eindeutig: Weniger erfahrene Personen wählen eher die Extremwerte.

**Tabelle 5-24: Signifikante Verteilungen der Bewertungskriterien innerhalb der Variable Erfahrung mehr oder weniger als ein Jahr (5er-Likert-Skala; mit \* markierte Zeilen, zeigen signifikante Zusammenhänge bei der berechneten 3er Likert-Skala)**

Rating	Exp.	$\chi^2$	Strongly disagree	disagree	Neither nor	agree	Strongly agree
<b>Likely solves the issue (K5)*</b>	<1 Jahr	$\chi^2(4)=9,921$ ,	9,2 %	13,3 %	24,5 %	42,9 %	10,2 %
	>1 Jahr	$p=,042$	10,7 %	21,8 %	30,5 %	29,8 %	7,2 %
<b>Elements I didn't think of (K3)*</b>	<1 Jahr	$\chi^2(4)=23,964$ ,	25,5 %	26,5 %	24,5 %	17,3 %	6,1 %
	>1 Jahr	$p<,001$	17,9 %	51 %	12,8 %	15,3 %	3,1 %
<b>Clearly stated (K8)*</b>	<1 Jahr	$\chi^2(4)=11,411$ ,	7,1 %	16,3 %	17,3 %	41,8 %	17,3 %
	>1 Jahr	$p=,022$	4 %	8,9 %	13,9 %	45,2 %	28,1 %
<b>Novel (K1)*</b>	<1 Jahr	$\chi^2(4)=14,598$ ,	11,2 %	22,4 %	17,3 %	39,8 %	9,2 %
	>1 Jahr	$p=,006$	11,3 %	34,2 %	24,3 %	24,1 %	6,1 %
<b>Unusual (K2)</b>	<1 Jahr	$\chi^2(4)=9,888$ ,	35,7 %	34,7 %	14,3 %	12,2 %	3,1 %
	>1 Jahr	$p=,042$	21,7 %	45,6 %	16,2 %	13,9 %	2,6 %

Die demografischen Daten zum Bildungsgrad der Probanden sind in dieser Teilstudie relativ gleich verteilt (Abbildung 5-14). Es zeigt sich hier, dass ein Großteil der Probanden ein Studium abgeschlossen hat. Die Personen ohne High-School-Abschluss werden in einem  $\chi^2$ -Test nicht berücksichtigt, da die Anzahl zu niedrig ist. Die Daten werden danach gruppiert, ob jemand Akademiker ist (Variable *Academics*).



**Abbildung 5-14: Bildungsgrad der Probanden in Teilstudie 2.3 (n=134)**

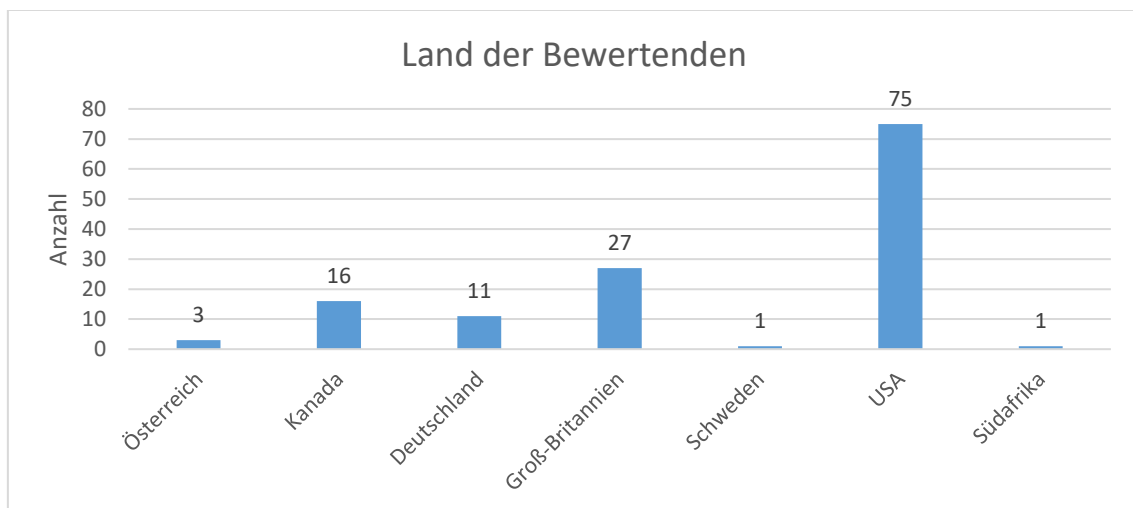
Ein  $\chi^2$ -Test mit der 5er-Likert-Skala hat gezeigt, dass der Bildungsgrad (Akademikern und Nicht-Akademikern) zu fast allen Bewertungskriterien einen signifikanten Zusammenhang aufweist (siehe Tabelle 9-9 in Anhang B). Die Verteilung der Werte ist ähnlich: Nicht-Akademiker bewerten im Vergleich zu Akademikern häufiger die beiden Extremwerte *strongly disagree* und *strongly agree*, wohingegen die Akademiker in diesen Fällen häufiger ein *disagree* und *agree* wählen. Die Vermutung ist, dass durch diese Konstellation die hohe Anzahl gefundener signifikanter Zusammenhänge zustande gekommen ist. Daher wurde für diese Analyse die 3er Likert-Skala genutzt. Für die reduzierte Skala wurden in einem  $\chi^2$ -Test weniger Zusammenhänge gefunden (siehe Tabelle 5-25).

Nicht-Akademiker bewerten Lösungsvorschläge häufig negativer als Akademiker. Nicht-Akademiker bewerten Lösungsvorschläge häufiger als neu (K1), sehen sie als seltener in die Praxis überführbar (K9), als schlechter beschrieben (K8) und seltener als ungewöhnlich an (K2). In Hinblick auf das Kriterium, wie sehr ein Lösungsvorschlag auf das Problem passt (K4), sind Akademiker häufiger unentschlossener.

**Tabelle 5-25: Zusammenhänge zwischen den Bewertungskriterien und der Variable *Academics* (3er-Likert-Skala)**

Rating	Academics	$\chi^2$ (df=2)	negative	Neither nor	positive
<b>Put into practice (K9)</b>	No Academic	$\chi^2=7,235$ ,	17,7 %	12,9 %	69,4 %
	Academic	$p=,027$	11,8 %	17,6 %	70,6 %
<b>Fits the issue (K4)</b>	No Academic	$\chi^2=9,747$ ,	15,9 %	10,8 %	73,4 %
	Academic	$p=,008$	12,9 %	18,6 %	68,5 %
<b>Clearly stated (K8)</b>	No Academic	$\chi^2=10,019$ ,	18,3 %	13,4 %	68,3 %
	Academic	$p=,007$	10,2 %	15,2 %	74,5 %
<b>Novel (K1)</b>	No Academic	$\chi^2=10,773$ ,	43,5 %	19,1 %	37,4 %
	Academic	$p=,005$	44,4 %	27,6 %	28,1 %
<b>Unusual (K2)</b>	No Academic	$\chi^2=11,138$ ,	72,3 %	11,6 %	16,1 %
	Academic	$p=,004$	63,3 %	20,2 %	16,5 %

In Abbildung 5-15 ist das Land der Crowdworker dargestellt. Zum einen fällt auf, dass es einer Person aus Südafrika gelungen ist, an der Teilstudie teilzunehmen, obwohl das Land innerhalb der Plattform nicht dafür freigeschaltet war. Da die Beiträge dieser Person vertrauenswürdig aussahen, wurden die Daten beibehalten. Zum anderen fällt auf, dass die überwiegende Mehrheit der Probanden Englisch als Muttersprache spricht (USA, Großbritannien und je nach Region auch Kanada). Hier werden die Länder in einer neuen Variable gruppiert, die unterscheidet, ob jemand Englisch als Muttersprache spricht (Variable *NativeSpeaker*). Kanada wird in diesem Fall zu den Ländern gezählt, in denen Englisch Muttersprache ist.



**Abbildung 5-15: Land der Probanden in Teilstudie 2.3 (n=134)**

Tabelle 5-26 zeigt die Verteilung der Bewertungen gruppiert nach den Personen, die Englisch als Muttersprache sprechen oder nicht sprechen. Hier zeigt sich insbesondere bei den Kategorien, wonach der Lösungsvorschlag einen Einfluss auf das eigene Verhalten haben kann (K10), dem Grad, wie sehr die Lösung auf das Problem passt (K4)



und ob der Lösungsvorschlag hilfreich ist (K6), dass die Personen, die Englisch nicht als Muttersprache haben, eher einen neutralen Wert in der Bewertung vergeben im Vergleich zu der anderen Gruppe. Hier kann es daran liegen, dass manche Lösungsvorschläge dahingehend nicht eindeutig formuliert sind, sodass Personen dieser Gruppe eher vorsichtig bewerten.

**Tabelle 5-26: Verteilung der Bewertung von Personen mit Englisch als Muttersprache (3er-Likert-Skala)**

<b>Rating</b>	<b>NativeSpeaker</b>	<b><math>\chi^2</math> (df=2)</b>	<b>negative</b>	<b>Neither nor</b>	<b>positive</b>
<b>Impact (K10)</b>	No Native Sp.	$\chi^2=7,990$ ,	18,8 %	31,3 %	50 %
	Native Sp.	$p=,018$	30,7 %	21,2 %	48,1 %
<b>Fits the issue (K4)</b>	No Native Sp.	$\chi^2=7,114$ ,	16,7 %	22,9 %	60,4 %
	Native Sp.	$p=,029$	14 %	13,5 %	72,5 %
<b>Elements I didn't think of (K3)</b>	No Native Sp.	$\chi^2=8,574$ ,	79,2 %	11,5 %	9,4 %
	Native Sp.	$p=,014$	64,8 %	14,8 %	20,4 %
<b>Helpful (K6)</b>	No Native Sp.	$\chi^2=7,321$ ,	12,5 %	27,1 %	60,4 %
	Native Sp.	$p=,026$	19,9 %	16,9 %	63,2 %

#### 5.3.4.2 Demografische Analyse der Autoren hinsichtlich der Lösungsvorschläge

Aus den vorherigen Studien existieren demografische Daten der Autoren der Lösungsvorschläge. In diesem Abschnitt werden mögliche Zusammenhänge zwischen den Bewertungskategorien und den demografischen Daten der Autoren untersucht.

Die statistische Grundlage sind  $\chi^2$ -Tests, die auf der gesamten Datenbasis durchgeführt werden. Problematisch ist, dass manche Ideen vier Mal und andere fünf Mal bewertet wurden, wodurch die demografischen Daten der Personen unterschiedlich oft einfließen. Gleichzeitig gab es in den Teilstudien 2.1 und 2.2 Beiträge mit mehr als einem Lösungsvorschlag, welche zwecks sauberer Bewertung in mehrere einzelne Beiträge aufgeteilt wurden. Dadurch entsteht ein Ungleichgewicht bei den demografischen Daten der Autoren. Würde man diese Daten allerdings nach Autoren gruppieren, so müssten die Durchschnittswerte der Bewertungen berechnet werden. Dadurch entsteht eine Vielzahl an kategorischen Variablen (die unterschiedlichen Dezimalzahlen), welche dann schwierig in  $\chi^2$ -Tests zu überprüfen sind, da es sehr viele Werte gibt und nicht alle eine ausreichende Anzahl für die Kalkulation haben.

Die  $\chi^2$ -Tests haben ergeben, dass es keinen Zusammenhang zwischen dem Alter bzw. dem Geschlecht und den Bewertungskategorien gibt.

Bei der Variable für Erfahrung zeigt sich, dass diese einen Zusammenhang mit dem Bewertungskriterium für die Klarheit der Formulierung hat (siehe Tabelle 5-27). Der  $\chi^2$ -Test zeigt, dass Personen mit mehr als einem Jahr Erfahrung eher Lösungsvorschläge formulieren, die als klar formuliert bewertet werden (K8).

**Tabelle 5-27: Zusammenhang zwischen der Variable ExpMoreThanAYear der Autoren und dem Kriterium für die Klarheit der Formulierung.**

Rating	Experience	$\chi^2$ (df=2)	negative	Neither nor	positive
<b>Clearly stated (K8)</b>	< 1 Jahr	$\chi^2=16,771$ ,	21,7 %	16,9 %	61,4 %
	> 1 Jahr	$p<,001$	11,4 %	13,4 %	75,3 %

Zwischen der Variable für den Bildungsgrad der Autoren und den Bewertungskategorien der Lösungsvorschläge gibt es ebenfalls mehrere Zusammenhänge (siehe Tabelle 5-28). Der Test zeigt, dass Autoren, die einen akademischen Abschluss haben (Bachelor, Master, Promotion) statistisch signifikant bessere Bewertungen in den Kategorien erhalten, wie sehr das Problem auf die Lösung passt (K4), inwieweit die Lösung das Problem lösen kann (K5), wie hilfreich der Lösungsvorschlag ist (K6), und ob der Lösungsvorschlag glaubwürdig ist (K7). Es zeigt sich, dass der Bildungsgrad einen höheren Einfluss auf die Qualität der Lösungsvorschläge hat als bspw. die Erfahrung des Autors.

**Tabelle 5-28: Zusammenhang zwischen der Variable Academics der Autoren und den Bewertungskriterien**

Rating	Academic	$\chi^2$ (df=2)	negative	Neither nor	positive
<b>Fits the issue (K4)</b>	No Academic	$\chi^2=9,227$	18,2 %	15,2 %	66,7 %
	Academic	$p=,010$	10,7 %	14,3 %	75 %
<b>Solves the issue (K5)</b>	No Academic	$\chi^2=7,661$ ,	35,8 %	29 %	35,2 %
	Academic	$p=,022$	36,8 %	30,5 %	42,7 %
<b>Helpful (K6)</b>	No Academic	$\chi^2=8,838$ ,	23,3 %	17,6 %	59,1 %
	Academic	$p=,012$	14,8 %	18,8 %	64,4 %
<b>Believable (K7)</b>	No Academic	$\chi^2=6,976$ ,	10,3 %	14,4 %	75,3 %
	Academic	$p=,031$	7,6 %	9,4 %	83,1 %

Die Muttersprache bzw. das Land, in dem die Autoren arbeiten, steht im Zusammenhang mit mehreren Bewertungskriterien, wie Tabelle 5-29 zeigt. Bei den folgenden Bewertungskriterien werden die Ideen von Personen, die Englisch als Muttersprache sprechen im Durchschnitt höher bewertet: Neuheit (K1), der Grad, wie sehr der Lösungsvorschlag zum Problem passt (K4), Glaubwürdigkeit (K7), Klarheit der Formulierung (K8) und wie sehr der Bewertende meint den Lösungsvorschlag des Autors in die Praxis überführen zu können (K9).

**Tabelle 5-29: Zusammenhang zwischen der Variable NativeSpeaker der Autoren und den Bewertungskriterien**

Rating	NativeSpeaker	$\chi^2$ (df=2)	negative	Neither nor	positive
<b>Novel (K1)</b>	No native Sp.	$\chi^2=9,425$ ,	51,6 %	24,7 %	23,6 %
	Native Speaker	$p=,009$	41,5 %	22,9 %	35,6 %
<b>Fits the issue (K4)</b>	No native Sp.	$\chi^2=6,953$ ,	18,1 %	18,7 %	63,2 %
	Native Speaker	$p=,031$	13,1 %	13,5 %	73,4 %
<b>Believable (K7)</b>	No native Sp.	$\chi^2=7,542$ ,	10,4 %	17 %	72,6 %
	Native Speaker	$p=,023$	8,4 %	10,2 %	81,4 %
<b>Clearly stated (K8)</b>	No native Sp.	$\chi^2=21,687$ ,	24,2 %	15,9 %	59,9 %
	Native Speaker	$p<,001$	11 %	13,8 %	75,1 %
<b>Put into practice (K9)</b>	No native Sp.	$\chi^2=8,761$ ,	20,9 %	17 %	62,1 %
	Native Speaker	$p=,013$	12,8 %	14,7 %	72,5 %

Größtenteils überlappen sich die Zusammenhänge zwischen den demografischen Daten der Autoren und den Bewertungskriterien nicht. Das Kriterium für die Glaubwürdigkeit (K7) hat Zusammenhänge sowohl zur Erfahrung als Crowdfunder als auch zum Land des Probanden. Der Grad, wie sehr der Lösungsvorschlag auf das Problem passt (K4), hat einen Bezug zum Bildungsgrad und zum Land des Autors.

#### 5.3.4.3 Unterschiede in den Probandengruppen zwischen Teilstudie 2.1 und 2.2

Es wurde ebenfalls evaluiert, inwieweit ein Unterschied zwischen den demografischen Daten der Probanden aus den beiden Teilstudien existiert. Dazu wurde ein Mann-Whitney-U-Test durchgeführt, dessen Ergebnis in Tabelle 5-30 abgebildet ist. Aus der Tabelle lässt sich entnehmen, dass insgesamt vier der demografischen Variablen signifikant unterschiedlich zwischen den beiden Gruppen sind: die Sprache bei den Bewertenden, und Alter, Geschlecht und die Bildung (Akademiker) bei den Autoren. Merkmale wie Anzahl der Wörter oder die aufgewendete Zeit sind in den Gruppen gleich verteilt.

Um mögliche Einflüsse der unterschiedlichen demografischen Variablen auf die Bewertungskategorien zu ermitteln, werden diese Zusammenhänge in Abschnitt 5.3.5.2 näher analysiert.

**Tabelle 5-30: Mann-Whitney-U Tests des Unterschiedes zwischen den Probandengruppen in den beiden Teilstudien 2.1 und 2.2. Es wurden nur Beiträge, die bei dem Lösungs-Prompt entstanden sind, berücksichtigt. Signifikante Unterschiede sind mit \* markiert.**

<b>Bewertende</b>	<b>U</b>	<b>Sig.</b>
Wochenarbeitsstunden	12225	,584
Muttersprache Englisch	11495	,021*
Erfahrung > 1 Jahr	12503	,728
Alter	11577	,294
Geschlecht	12390	,769
Akademiker (Bildung)	12357	,675
<b>Autoren</b>	<b>U</b>	<b>Sig.</b>
Wochenarbeitsstunden	12051,5	,425
Muttersprache Englisch	11680	,134
Erfahrung > 1 Jahr	12540,5	,846
Alter	10922,5	,032*
Geschlecht	11020,5	,024*
Akademiker (Bildung)	8834	<,001*
<b>Allgemeines</b>	<b>U</b>	<b>Sig.</b>
Zeit	11345,5	,507
Anzahl der Wörter	12603	,943

### 5.3.5 Analyse der Daten

In diesem Abschnitt werden die Bewertungskriterien mit unterschiedlichen Perspektiven analysiert. Zunächst erfolgt eine Betrachtung der Bewertungen von Lösungsvorschlägen anhand der beiden Konditionen mit Diskussion und ohne Diskussion. Anschließend erfolgt eine Betrachtung, ob es Zusammenhänge zwischen den Bewertungskriterien und den reflexionsbezogenen Codes gibt. Eine Beschreibung der Vorgehensweise für die statistische Analyse, sowie die Überprüfung der Daten auf eventuelle Voraussetzungen für statistische Verfahren ist in Anhang C.

#### 5.3.5.1 Analyse der Lösungsvorschläge von Teilstudie 2.1 und 2.2

Das Ziel der Teilstudie ist der Vergleich der Lösungsvorschläge, die nach Prompts in Themas mit und ohne Diskussion erstellt wurden. Dafür werden zunächst die dafür erstellten Kriterien analysiert.

Für den Großteil der Bewertungskategorien liegt die Standardabweichung knapp über oder unter eins (Tabelle 5-31). Das impliziert, dass die überwiegende Mehrheit der Bewertenden sich auf eine Bewertung einigen konnte.

**Tabelle 5-31: Durchschnitt und Standardabweichung der Bewertungen je Kategorie und Studie**

	Teilstudie 2.1 (n=369)		Teilstudie 2.2 (n=384)		Insgesamt (n=753)	
	Mean	SD	Mean	SD	Mean	SD
Novel (K1)	2,89	1,148	2,79	1,111	2,84	1,130
Unusual (K2)	2,21	1,068	2,34	1,032	2,28	1,051
Elements I didn't think of (K3)	2,38	1,112	2,36	1,015	2,37	1,063
Fits the issue (K4)	3,82	0,965	3,69	1,087	3,75	1,030
Likely solves the issue (K5)	3,11	1,087	2,99	1,140	3,05	1,115
Helpful (K6)	3,62	1,085	3,56	1,134	3,59	1,110
Believable (K7)	3,99	0,895	3,90	0,941	3,94	0,919
Clearly stated (K8)	3,87	0,991	3,72	1,148	3,80	1,076
Put into practice (K9)	3,82	1,073	3,68	1,109	3,75	1,093
Impact (K10)	3,23	1,155	3,17	1,171	3,20	1,163
Recommend (K11)	3,45	1,204	3,32	1,271	3,38	1,240

Für den Vergleich der Bewertungskategorien wird ein Mann-Whitney-U-Test durchgeführt (siehe Abschnitt C) bei der sowohl die Bewertungen der 5er-Likert-Skala als auch die berechnete 3er-Likert-Skala (siehe Abschnitt 5.3.4) miteinander vergleicht. Die Ergebnisse dieser Berechnung für alle Prompts finden sich in Tabelle 5-32.

Zunächst werden die Ergebnisse der 5er-Likert-Skala betrachtet. Die Tabelle 5-32 zeigt einen signifikanten Unterschied in den Bewertungen zwischen den beiden Konditionen für die Bewertungskategorie, ob der Vorschlag ungewöhnlich ist. Hier gab es in der Teilstudie 2.2, in der das Thema Diskussionsbeiträge hatte, signifikant höhere Werte, also mehr Zustimmung. Dies bedeutet, dass durch Prompts hervorgerufene erfahrungsbasierte Lösungsvorschläge innerhalb einer Diskussion eher ungewöhnlich sind als zu Beginn einer Diskussion. Die Bewertungskategorie, ob der Lösungsvorschlag leicht in die Praxis überführt werden könnte, ist knapp nicht signifikant, wobei die Lösungsvorschläge in dem Thema ohne Antworten bessere Bewertungen bekommen haben. Gemäß allen anderen Kategorien gibt es keinen Unterschied, inwieweit ein Lösungsvorschlag neu ist (K1), Elemente enthält, an die der Bewertende nicht gedacht hat (K3), die Lösung auf das Problem passt (K4), der Vorschlag das Problem löst (K5), oder der Vorschlag hilfreich (K6), glaubwürdig (K7) oder klar verständlich ist (K8). Es gibt ebenfalls keinen Unterschied hinsichtlich des Grads, wie sehr der Lösungsvorschlag die Praxis des Bewertenden beeinflusst (K10) bzw. wie sehr derjenige den Vorschlag weiterempfehlen würde (K11).

Innerhalb der 3er-Likert-Skala ist für die Bewertungskategorie für ungewöhnliche Inhalte (K2) die Bewertung in der Kondition mit Diskussionsbeiträgen ebenfalls höher, allerdings ist dieser Unterschied nicht signifikant. Bei drei Bewertungskategorien, dem

Grad wie sehr die Lösung auf das Problem passt (K4), die Klarheit der Formulierung (K8) und ob der Bewertende sich in der Lage fühlt, die Lösung in die Praxis umzusetzen (K9), findet sich ein signifikanter Unterschied zwischen den beiden Konditionen. Bei allen drei Bewertungskategorien waren die Bewertungen für die Kondition ohne Diskussionsbeiträge höher. Bei den übrigen Bewertungskategorien gibt es keine signifikanten Unterschiede zwischen den Konditionen (Neuheit, Elemente, an die der Bewertende nicht gedacht hat, ob der Vorschlag das Problem löst, wie hilfreich oder glaubwürdig der Vorschlag formuliert ist, der mögliche Einfluss auf die Handlungen des Bewertenden und ob dieser den Vorschlag weiterempfehlen würden).

**Tabelle 5-32: Mann-Whitney-U Test der (\*signifikanten) Bewertungskategorien zwischen den Konditionen mit und ohne Diskussionen (alle Prompts)**

	5er-Likert-Skala		3er-Likert-Skala	
	U	Sig.	U	Sig.
Novel (K1)	67238	,210	67661,	,251
Unusual (K2)	65171,5	,044*	66888,	,108
Elements I didn't think of (K3)	70419,5	,878	68938,5	,442
Fits the issue (K4)	66913	,157	66117,	,047*
Likely solves the issue (K5)	66827,5	,163	67410,5	,220
Helpful (K6)	68914	,496	67913,	,253
Believable (K7)	67321,5	,194	67652,	,130
Clearly stated (K8)	67090	,181	65519,	,024*
Put into practice (K9)	65490	,057	66051,	,046*
Impact (K10)	68659	,446	68975,5	,496
Recommend (K11)	66582,5	,137	66547,5	,104
Use it (K12)			67252,5	,188

Bei der für Tabelle 5-32 durchgeführten Diskussion gab es im Vergleich allerdings nicht nur den Aufbau des Themas bzw. der Diskussionsbeiträge als Unterschied, sondern auch, dass sowohl die erfahrungsbasierten Lösungen der Lösungs-Prompts und der Kontroll- (Lösungs-) Prompts bzw. Erfahrungs-Prompts eingeflossen sind. Um diesen Einfluss zu minimieren, wurde die gleiche Berechnung durchgeführt, wobei der Datensatz ausschließlich auf die Lösungs-Prompts limitiert war. Das Ergebnis findet sich in Tabelle 5-33. Zunächst zeigt die Tabelle, dass es für die Kategorien, ob ein Vorschlag neu (K1) oder ungewöhnlich ist (K2), oder ob er Elemente enthält, an die der Bewertende nicht gedacht hat (K3), keinen signifikanten Unterschied zwischen den beiden Konditionen gibt. Bei allen anderen Kategorien gibt es einen signifikanten Unterschied, der für alle Bewertungskategorien so ausfällt, dass die Lösungsvorschläge aus der Kondition ohne Diskussionsbeiträge höhere Bewertungen erzielen als die Bewertungen mit Diskussionsbeiträgen.

**Tabelle 5-33: Mann-Whitney-U Test der (\*signifikanten) Bewertungskategorien für Vorschläge aus Lösungs-Prompts zwischen den Konditionen mit (M) und ohne (O) Diskussionen (n=331)**

	5er-Likert-Skala		3er-Likert-Skala		O/M
	U	Sig.	U	Sig.	
Novel (K1)	11584,5	,183	11619,5	,181	
Unusual (K2)	11292	,082	11664	,147	
Elements I didn't think of (K3)	12206,5	,559	12221,5	,521	
Fits the issue (K4)	10285	,002*	10075	<,001*	O > M
Likely solves the issue (K5)	9927,5	,001*	9998	,001*	O > M
Helpful (K6)	10776,5	,018*	10671,5	,005*	O > M
Believable (K7)	10867	,019*	11321,5	,026*	O > M
Clearly stated (K8)	10941	,030*	10897,5	,010*	O > M
Put into practice (K9)	10626	,010*	10686	,003*	O > M
Impact (K10)	10855,5	,025*	10995	,031*	O > M
Recommend (K11)	10521,5	,008*	10598	,005*	O > M
Use it (K12)			10609	,007*	O > M

Ein Kriterium bei dem Vergleich der Lösungsvorschläge ist die Anzahl der Wörter (K13). Eine ANOVA kommt nicht infrage, da sowohl die Gruppen aus Teilstudie 2.1 und 2.2 nicht gleich groß sind, als auch die Homogenität der Varianzen nicht erfüllt ist (Field 2009, 360). Hierfür wurden in einem Mann-Whitney-U-Test die gruppierten Anzahlen der Wörter verglichen. Das Ergebnis ist, dass es keinen Unterschied zwischen den Konditionen gibt und somit die Prompts nicht mehr oder weniger Wörtern führen. Dieses Ergebnis findet sich sowohl unter Einbezug aller Prompts und auch, wenn nur die Ergebnisse für den Lösungs-Prompt einbezogen werden.

Um die Größe des Unterschieds der Bewertung besser einschätzen zu können, wurden die durchschnittlichen Bewertungen jedes Kriteriums für den Lösungs-Prompt in Tabelle 5-34 dargestellt. Es zeigt sich, dass die Bewertung für einen Lösungsvorschlag aus Teilstudie 2.1 um durchschnittlich 0,26 höher ist als ein Lösungsvorschlag aus Teilstudie 2.2. Die Standardabweichung ist in Teilstudie 2.2 um durchschnittlich 0,13 höher, was bedeutet, dass die Werte ganz leicht mehr streuen. Somit ist ein Unterschied zwischen den Bewertungen vorhanden, wenn er in der Realität ca. ein Viertel einer Bewertungsstufe ausmacht (wobei die Bewertungsstufen nicht auf einer äquidistanten Skala sind).

**Tabelle 5-34: Vergleich der Bewertungen für den Lösungs-Prompt  
zwischen Teilstudie 2.1 und 2.2 (5er-Likert-Skala, außer 3er-Likert-Skala bei K12)**

	Teilstudie 2.1 (ohne Diskussion)		Teilstudie 2.2 (mit Diskussion)	
	Mittelwert	St. Abw.	Mittelwert	St. Abw.
Novel (K1)	2,88	1,146	2,71	1,111
Unusual (K2)	2,17	1,040	2,35	1,028
Elements I didn't think of (K3)	2,34	1,000	2,29	1,013
Fits the issue (K4)	3,98	0,840	3,56	1,171
Likely solves the issue (K5)	3,35	0,993	2,91	1,168
Helpful (K6)	3,83	0,990	3,50	1,189
Believable (K7)	4,08	0,816	3,82	0,985
Clearly stated (K8)	3,95	0,942	3,65	1,151
Put into practice (K9)	4,00	0,987	3,66	1,169
Impact (K10)	3,43	1,067	3,13	1,190
Recommend (K11)	3,65	1,142	3,26	1,300
Use it (K12)	2,45	0,787	2,20	0,840

### 5.3.5.2 Regressionsmodelle zur Überprüfung des Einflusses demografischer Variablen

Um zu überprüfen, inwieweit die demografischen Daten bei dem im vorherigen Abschnitt beobachteten Ergebnis eine Rolle spielen, wird eine Regressionsanalyse durchgeführt. Hierbei wird eine ordinale Regression angewendet, um zu überprüfen, welche Variable zur Varianz der Bewertungskategorie beiträgt. Die dafür benötigten Grundlagen und die Überprüfung der Voraussetzungen sind in Anhang C dargestellt.

Wie Tabelle 9-5 in Anhang C zeigt, sind bei der 5er-Likert-Skala gültige Regressionsmodelle für die Bewertungskategorien für den Grad der Glaubwürdigkeit (K7) und inwieweit jemand den Lösungsvorschlag weiterempfehlen würde (K11) entstanden. Zusätzlich hat die Bewertungskategorie, die abbildet, inwieweit der Bewertende den Lösungsvorschlag nutzen würde, sofern er das Problem hätte (K12) ebenfalls ein gültiges Regressionsmodell. Diese Kategorie wurde in einer 3er-Likert-Skala erhoben.

Um ein ordinales Regressionsmodell zu erhalten und weiterhin ein Modell mit allen demografischen Variablen zu behalten, wurde ebenfalls Regressionsmodelle auf Basis der reduzierten 3er-Likert-Skala berechnet (siehe Anhang C Tabelle 9-7). Der Vorteil eines ordinalen Regressionsmodells ist, dass die gefundenen Zusammenhänge für jede Ausprägung des Merkmals der abhängigen Variable gültig sind. Bei den



Regressionsmodellen auf Basis der 3er-Likert-Skala wurden die obigen Regressionsmodelle ebenfalls gefunden. Zusätzlich gab es gültige Regressionsmodelle für die Variablen für die Wahrscheinlichkeit, wie sehr der Lösungsvorschlag das Problem lösen kann (K5), der klaren Beschreibung des Lösungsvorschlags (K8) und der Auswirkung auf das künftige Verhalten des Bewertenden (K10).

Für die Bewertungskategorien der Neuheit (K1) und der Elemente, an die Bewertende nicht gedacht hat (K3) gab es für beide Likert-Skalen keine Regressionsmodelle (siehe Anhang C Tabelle 9-5 und Tabelle 9-7), bei denen sich ein Modell mit den unabhängigen Variablen stark von einem Modell ohne die unabhängigen Variablen unterscheidet. Dementsprechend kann hier die Varianz der Variablen nicht erklärt werden.

Für die übrigen Bewertungskategorien (K4 und K6) gibt es kein gültiges ordinales Regressionsmodell (siehe Anhang C Tabelle 9-5 und Tabelle 9-7). Hier sind weitere Analysen notwendig, um den Einfluss der demografischen Variablen zu überprüfen.

In den folgenden Abschnitten werden die signifikanten Prädiktoren und ihr Einfluss auf die jeweilige Bewertungskategorie (außer K1, K3, K4, K6) dargestellt. Es erfolgt dabei der Übersichtlichkeit halber eine Trennung, bei der das vollständige Modell der ordinalen Regression in Anhang 0 dargestellt ist und die Wahrscheinlichkeiten für einzelne Prädiktoren in den folgenden Abschnitten dargestellt werden. Dabei werden nur die unabhängigen Variablen berücksichtigt, bei denen jede Ausprägung einen signifikanten Zusammenhang hatte (bspw. jede Kategorie in der gruppierten Variable für das Alter).

Die folgenden Modelle dürfen nur vorsichtig interpretiert werden, da aufgrund der Anzahl der Variablen in den Regressionsmodellen viele Ausprägungskombinationen (ca. 65 %) fehlen. In der Literatur werden solche und höhere Werte allerdings noch genutzt, was in Anhang C konkreter erläutert wird.

#### 5.3.5.2.1 Ordinale Regressionsanalyse für die Bewertungskategorie für ungewöhnliche Elemente (K2)

Um zu prüfen, zu welchem Grad ein Lösungsvorschlag ungewöhnliche Elemente enthält, wurde ein valides ordinales Regressionsmodell für die zusammengefasste 3er-Likert-Skala berechnet (Tabelle 9-7 in Anhang C). Im Regressionsmodell (siehe Tabelle 9-12 in Anhang 0) sind die Prädiktoren Zeit, Englisch als Muttersprache (Autor), mehr Erfahrung als 1 Jahr (Autor), Wochenstunden (Autor) und Alter (Autor). In der folgenden Tabelle 5-35 sind die Wahrscheinlichkeiten für die Ausprägungen der signifikanten unabhängigen Variablen abgebildet.

Zunächst zeigt sich direkt, dass das Vorhandensein von Diskussionsbeiträgen (der Unterschied zwischen den Teilstudien 2.1 und 2.2; Variable im Folgenden als *Studie* bezeichnet) kein signifikanter Faktor in diesem Regressionsmodell ist. Zunächst ist an den Ergebnissen auffällig, dass Personen mit weniger Erfahrung eher besser bewertete Beiträge verfassen als Personen mit mehr als einem Jahr Erfahrung. Ebenfalls haben Personen, die nicht Englisch als Muttersprache sprechen ebenso eine höhere Wahrscheinlichkeit auf höher bewertete Beiträge. Bei beiden Variablen sind jeweils weniger Personen in dieser Gruppe, was einen statistischen Effekt begünstigt haben kann. Bei der aufgewendeten Zeit zeigt sich, dass die geringste Wahrscheinlichkeit auf eine positive Bewertung bei der wenigsten aufgewendeten Zeit liegt. Hier steigt die Wahrscheinlichkeit mit mehr aufgewendeter Zeit aber nicht linear! Dies bedeutet, dass Lösungsvorschläge mit ungewöhnlichen Elementen auch in kurzer Zeit zu einer Diskussion beigetragen werden können. Hinsichtlich des Alters ergibt sich kein klares Bild und sowohl die Autoren in den Altersgruppen 18-24 und 55-64 haben eine höhere Wahrscheinlichkeit auf Lösungsvorschläge mit ungewöhnlichen Elementen. Hier kann allerdings auch die niedrige Anzahl der Grund für einen statistischen Effekt sein. Bei der Wochenarbeitszeit ist das Bild ebenfalls nicht klar. Es lässt sich aber ableiten, dass Personen, die besonders viel als Crowdworker arbeiten (> 30 Stunden pro Woche) eher weniger einen Lösungsvorschlag als ungewöhnlich bewerten.

**Tabelle 5-35: Ordinales Regressionsmodell für das Kriterium für ungewöhnliche Elemente in einem Lösungsvorschlag (K2). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird.**

		<b>Zeit</b>		
<b>Zeit</b>		negativ	neutral	positiv
<b>0-240s</b> (n=28)	Mean	76,17%	12,72%	11,11%
	Std. Deviation	0,15469	0,07284	0,08294
<b>241-480s</b> (n=139)	Mean	65,07%	16,57%	18,36%
	Std. Deviation	0,18637	0,05754	0,13964
<b>481-720s</b> (n=83)	Mean	72,52%	14,60%	12,88%
	Std. Deviation	0,13549	0,05897	0,07877
<b>720s+</b> (n=76)	Mean	66,42%	16,63%	16,95%
	Std. Deviation	0,16938	0,06340	0,10921
<b>Autor: Muttersprache Englisch</b>				
<b>Autor: Muttersprache Englisch</b>		negativ	neutral	positiv
<b>No NativeSpeaker</b> (n=92)	Mean	64,56%	17,19%	18,24%
	Std. Deviation	0,17758	0,06585	0,11551
<b>NativeSpeaker</b> (n=234)	Mean	69,68%	15,18%	15,13%
	Std. Deviation	0,16763	0,05923	0,11782

<b>Autor: Erfahrung &gt; 1 Jahr</b>				
<b>Autor: Erfahrung &gt; 1 Jahr</b>		negativ	neutral	positiv
<b>&lt; 1 Jahr Erfahrung</b>	Mean	60,08%	17,88%	22,04%
(n=76)	Std. Deviation	0,20182	0,05915	0,15833
<b>&gt; 1 Jahr Erfahrung</b>	Mean	70,72%	15,10%	14,18%
(n=250)	Std. Deviation	0,15361	0,06116	0,09553
<b>Autor: Wochenarbeitsstunden</b>				
<b>Autor: Wochenarbeitsstunden</b>		negativ	neutral	positiv
<b>Weniger als 1 Stunde / Woche</b>	Mean	70,94%	15,88%	13,18%
(n=10)	Std. Deviation	0,07860	0,03272	0,04623
<b>1 - 5 Stunden / Woche</b>	Mean	64,48%	16,76%	18,75%
(n=170)	Std. Deviation	0,19125	0,06336	0,13801
<b>5 - 15 Stunden / Woche</b>	Mean	72,90%	14,48%	12,62%
(n=93)	Std. Deviation	0,12889	0,05444	0,07679
<b>15 - 30 Stunden / Woche</b>	Mean	68,97%	15,89%	15,14%
(n=38)	Std. Deviation	0,15295	0,06022	0,09575
<b>&gt; 30 Stunden / Woche</b>	Mean	78,19%	11,73%	10,09%
(n=15)	Std. Deviation	0,16016	0,07883	0,08187
<b>Autor: Alter</b>				
<b>Autor: Alter</b>		negativ	neutral	positiv
<b>Kein Alter angegeben</b>	Mean	89,74%	6,25%	4,01%
(n=10)	Std. Deviation	0,03435	0,02010	0,01426
<b>18-24</b>	Mean	65,41%	16,38%	18,20%
(n=33)	Std. Deviation	0,19042	0,05858	0,14386
<b>25-34</b>	Mean	71,37%	14,33%	14,30%
(n=116)	Std. Deviation	0,17161	0,05818	0,12495
<b>35-44</b>	Mean	66,35%	17,00%	16,65%
(n=67)	Std. Deviation	0,15085	0,05940	0,09471
<b>45-54</b>	Mean	67,89%	16,26%	15,85%
(n=59)	Std. Deviation	0,15406	0,05578	0,10192
<b>55-64</b>	Mean	59,98%	18,83%	21,18%
(n=41)	Std. Deviation	0,17660	0,06262	0,11747

#### 5.3.5.2.2 Ordinale Regressionsanalyse für die Bewertungskategorie für den Grad wie sehr der Lösungsvorschlag das Problem lösen kann (K5)

Die Validität des Regressionsmodells für den Grad inwieweit der Lösungsvorschlag das Problem lösen kann (K5) ist in Tabelle 9-7 in Anhang C dargestellt (3er-Likert-Skala). Im Regressionsmodell (siehe Tabelle 9-13 in Anhang 0) ist mindestens eine Ausprägung der folgenden Prädiktoren signifikant: Anzahl der Wörter, die Studie, Erfahrung mehr als 1 Jahr (Autor), Alter (Autor) und Akademiker (Autor). Mit Studie ist der Unterschied in der Bewertung der Lösungsvorschläge zwischen Teilstudie 2.1 und 2.2 gemeint. In Tabelle 5-36 sind die Wahrscheinlichkeiten für die Ausprägungen der signifikanten unabhängigen Variablen abgebildet.

**Tabelle 5-36: Ordinales Regressionsmodell für das Kriterium den Grad wie sehr der Lösungsvorschlag das Problem lösen kann (K5). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird.**

<b>Anzahl der Wörter</b>				
Anzahl der Wörter		negativ	neutral	positiv
<b>0-40</b>	Mean	86,35%	9,77%	3,87%
(n=9)	Std. Deviation	0,13072	0,09166	0,03921
<b>41-80</b>	Mean	33,60%	30,41%	35,98%
(n=116)	Std. Deviation	0,18359	0,05749	0,17905
<b>81-120</b>	Mean	28,78%	30,62%	40,60%
(n=107)	Std. Deviation	0,15230	0,07033	0,18648
<b>121+</b>	Mean	24,52%	29,77%	45,71%
(n=94)	Std. Deviation	0,13888	0,05403	0,17954
<b>Studie</b>				
Studie		negativ	neutral	positiv
<b>Teilstudie 2.1</b>	Mean	21,72%	28,31%	49,97%
(n=120)	Std. Deviation	0,15431	0,06186	0,17330
<b>Teilstudie 2.2</b>	Mean	36,18%	30,55%	33,27%
(n=206)	Std. Deviation	0,18649	0,07398	0,17607
<b>Autor: Erfahrung &gt; 1 Jahr</b>				
Autor: Erfahrung > 1 Jahr		negativ	neutral	positiv
<b>&lt; 1 Jahr Erfahrung</b>	Mean	40,41%	28,64%	30,95%
(n=76)	Std. Deviation	0,23792	0,08779	0,18121
<b>&gt; 1 Jahr Erfahrung</b>	Mean	27,96%	30,05%	41,99%
(n=250)	Std. Deviation	0,16031	0,06419	0,18875
<b>Autor: Alter</b>				
Autor: Alter		negativ	neutral	positiv
<b>Kein Alter angegeben</b>	Mean	24,94%	31,61%	43,45%
(n=10)	Std. Deviation	0,11110	0,04312	0,15033
<b>18-24</b>	Mean	18,90%	27,04%	54,05%
(n=33)	Std. Deviation	0,11812	0,09295	0,20110
<b>25-34</b>	Mean	34,10%	27,40%	38,50%
(n=116)	Std. Deviation	0,24475	0,08445	0,22044
<b>35-44</b>	Mean	32,86%	30,89%	36,25%
(n=67)	Std. Deviation	0,17065	0,05502	0,17375
<b>45-54</b>	Mean	34,12%	33,19%	32,69%
(n=59)	Std. Deviation	0,11223	0,03333	0,12532
<b>55-64</b>	Mean	24,79%	31,11%	44,09%
(n=41)	Std. Deviation	0,12050	0,03874	0,14983
<b>Autor: Akademiker (Bildung)</b>				
Autor: Akademiker (Bildung)		negativ	neutral	positiv
<b>No Academic</b>	Mean	34,40%	31,52%	34,07%
(n=141)	Std. Deviation	0,16265	0,05039	0,15612
<b>Academic</b>	Mean	28,16%	28,35%	43,49%
(n=185)	Std. Deviation	0,20233	0,08006	0,20750

Hinsichtlich dieses Kriteriums haben die demografischen Variablen der Bewertenden keinen Einfluss auf die Bewertung. In diesem Kriterium hat die Variable für die Studie einen Einfluss. Die Wahrscheinlichkeit auf eine positive Bewertung steigt um ca. 17 %, wenn der Autor keine Diskussion sieht und direkt einen Lösungsvorschlag verfasst. Bei der Anzahl der Wörter ist für dieses Kriterium eine stetige Steigerung der Wahrscheinlichkeit für eine positive Bewertung vorhanden. Je mehr Wörter verfasst werden, desto eher wird ein Lösungsvorschlag hinsichtlich des Kriteriums, wie sehr der Lösungsvorschlag das Problem löst, positiv bewertet. Bei den demografischen Daten des Autors ist die Bildung, das Alter und die Erfahrung relevant. Wenn der Autor erfahren ist (>1 Jahr Erfahrung) oder wenn er studiert hat (Akademiker als Bildung), dann ist die Wahrscheinlichkeit 9 % bzw. 11 % höher einen positiv bewerteten Lösungsvorschlag in dieser Kategorie zu erhalten (im Vgl. zu der jeweiligen anderen Ausprägung des Merkmals). Hinsichtlich des Alters scheinen jüngere (18-24) oder ältere (55-64) Personen am ehesten positiv bewertete Lösungsvorschläge beizutragen. 10 Personen haben ihr Alter nicht angegeben. Aufgrund der niedrigen Anzahl sollte der hohe Anteil positiv bewerteter Lösungsvorschläge nicht überbewertet werden. Die größten Einflüsse auf die Wahrscheinlichkeit, eine positive Bewertung zu erhalten, haben bei dieser Bewertungskategorie die Verwendung von mehr als 40 Wörtern in der Antwort und das Fehlen von übrigen Diskussionsbeiträgen.

#### 5.3.5.2.3 Ordinale Regressionsanalyse für die Bewertungskategorie für die Glaubwürdigkeit (K7)

Für die Variable inwieweit der Lösungsvorschlag glaubwürdig ist, wurde ein gültiges ordinales Regressionsmodell für die 5er-Likert-Skala berechnet (Tabelle 9-5 in Anhang C). Im Regressionsmodell (Tabelle 9-14 in Anhang 0) wurde ermittelt, dass die Variablen für das Alter des Bewertenden, die Studie und die Anzahl der Wörter signifikant sind. Zusätzlich sind beim Autor die Variablen für Alter, Akademiker und die Erfahrung signifikante Prädiktoren. In Tabelle 5-37 sind die Wahrscheinlichkeiten für die Ausprägungen der signifikanten unabhängigen Variablen abgebildet.

**Tabelle 5-37: Ordinales Regressionsmodell für das Kriterium den Grad der Glaubwürdigkeit (K7).**  
**Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an,**  
**mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird.**

		<b>Bewertender: Alter</b>				
Bewertender: Alter		strongly disagree	disagree	neither nor	agree	strongly agree
<b>17-24</b>	Mean	0,93%	4,70%	8,22%	52,11%	34,03%
(n=23)	Std. Deviation	0,00852	0,03910	0,05474	0,08213	0,16317
<b>25-34</b>	Mean	2,34%	8,90%	12,75%	53,87%	22,14%
(n=86)	Std. Deviation	0,04749	0,07291	0,06594	0,08652	0,14114
<b>35-44</b>	Mean	2,72%	10,23%	14,34%	54,81%	17,90%
(n=76)	Std. Deviation	0,05139	0,07307	0,06143	0,08586	0,10395
<b>45-54</b>	Mean	2,09%	8,47%	12,52%	54,95%	21,97%
(n=71)	Std. Deviation	0,03745	0,06885	0,06076	0,07780	0,12844
<b>55+</b>	Mean	1,25%	4,93%	7,82%	51,84%	34,17%
(n=70)	Std. Deviation	0,03317	0,06088	0,04995	0,09096	0,14918
		<b>Anzahl der Wörter</b>				
Anzahl der Wörter		strongly disagree	disagree	neither nor	agree	strongly agree
<b>0-40</b>	Mean	17,86%	26,22%	16,38%	33,73%	5,81%
(n=9)	Std. Deviation	0,18361	0,15286	0,04530	0,24553	0,05374
<b>41-80</b>	Mean	1,78%	8,07%	12,07%	54,40%	23,68%
(n=116)	Std. Deviation	0,01875	0,06709	0,06601	0,06997	0,13648
<b>81-120</b>	Mean	1,83%	8,56%	13,05%	54,69%	21,87%
(n=107)	Std. Deviation	0,01470	0,05908	0,06482	0,07062	0,14301
<b>121+</b>	Mean	1,09%	5,42%	9,23%	53,84%	30,41%
(n=94)	Std. Deviation	0,00983	0,04297	0,05571	0,06829	0,14697
		<b>Studie</b>				
Studie		strongly disagree	disagree	neither nor	agree	strongly agree
<b>Teilstudie 2.1</b>	Mean	1,09%	5,36%	9,09%	54,24%	30,22%
(n=120)	Std. Deviation	0,01081	0,04492	0,05380	0,06557	0,13664
<b>Teilstudie 2.2</b>	Mean	2,59%	9,49%	13,21%	53,49%	21,22%
(n=206)	Std. Deviation	0,05160	0,07782	0,06539	0,09543	0,14366
		<b>Autor: Erfahrung &gt; 1 Jahr</b>				
Autor: Erfahrung > 1 Jahr		strongly disagree	disagree	neither nor	agree	strongly agree
<b>&lt; 1 Jahr Erfahrung</b>	Mean	4,15%	11,99%	14,36%	51,71%	17,79%
(n=76)	Std. Deviation	0,08126	0,10253	0,06611	0,12747	0,12225
<b>&gt; 1 Jahr Erfahrung</b>	Mean	1,40%	6,74%	10,88%	54,39%	26,59%
(n=250)	Std. Deviation	0,01228	0,05156	0,06182	0,06714	0,14856

		<b>Autor: Alter</b>				
Autor: Alter		strongly disagree	disagree	neither nor	agree	strongly agree
<b>Kein Alter angegeben</b> (n=10)	Mean	0,77%	4,04%	7,70%	55,97%	31,53%
	Std. Deviation	0,00364	0,01814	0,02976	0,04993	0,09766
<b>18-24</b> (n=33)	Mean	0,67%	3,50%	6,65%	51,54%	37,64%
	Std. Deviation	0,00470	0,02306	0,03709	0,09219	0,14730
<b>25-34</b> (n=116)	Mean	3,26%	10,45%	13,46%	52,17%	20,65%
	Std. Deviation	0,06716	0,09151	0,06853	0,11137	0,14569
<b>35-44</b> (n=67)	Mean	1,65%	7,63%	11,51%	53,11%	26,10%
	Std. Deviation	0,01646	0,06541	0,07097	0,06809	0,15590
<b>45-54</b> (n=59)	Mean	1,51%	7,47%	12,40%	57,39%	21,23%
	Std. Deviation	0,00863	0,03849	0,05001	0,03463	0,10494
<b>55-64</b> (n=41)	Mean	1,39%	6,75%	11,00%	55,38%	25,48%
	Std. Deviation	0,01185	0,04964	0,05890	0,06019	0,13487
		<b>Autor: Akademiker (Bildung)</b>				
Autor: Akademiker (Bildung)		strongly disagree	disagree	neither nor	agree	strongly agree
<b>No Academic</b> (n=141)	Mean	2,17%	9,83%	14,35%	54,77%	18,89%
	Std. Deviation	0,01795	0,06549	0,06410	0,06920	0,12179
<b>Academic</b> (n=185)	Mean	1,94%	6,55%	9,67%	53,00%	28,84%
	Std. Deviation	0,05374	0,07087	0,05708	0,09576	0,15103

Hinsichtlich der Studie zeigt diese Analyse, dass das Vorhandensein von Diskussionsbeiträgen die Wahrscheinlichkeit auf eine sehr hohe Zustimmung bei der Bewertung, dass die Lösung glaubwürdig ist, um 9 % senkt. Die Nutzung von 40 oder weniger Wörter hat die niedrigste Wahrscheinlichkeit auf positive Bewertungen, während die Wahrscheinlichkeit bei 120 und mehr Wörtern leicht höher ist als bei den anderen Ausprägungen. Hier zeigt sich, dass sehr kurze Beiträge keine hohe Chance haben, Glaubwürdigkeit zu vermitteln. In dieser Studie hat das Alter des Bewertenden einen Einfluss auf die Wahrscheinlichkeit, dass ein Lösungsvorschlag als glaubwürdig bewertet wird. In der Altersgruppe, die in der Gruppierung in der Mitte ist (35-44), ist die Wahrscheinlichkeit am niedrigsten, während die Randgruppen (17-24 und 55+) die höchste Wahrscheinlichkeit auf eine positive Bewertung haben. Beim Autor sind die demografischen Daten für Alter, Erfahrung und Bildung ein signifikanter Prädiktor. Wenn der Autor besonders erfahren ist (> 1 Jahr) oder studiert hat (Akademiker), dann steigt die Wahrscheinlichkeit auf eine positive Bewertung hinsichtlich der Glaubwürdigkeit. Beim Alter ist die Verteilung grob ähnlich der Verteilung der Ausprägungen des Alters der Bewertenden: Die Altersgruppe für jüngere (18-24) und ältere (55+) Autoren bewerten positiver. Hier sind die Unterschiede zwischen den Gruppen aber eher gering. In dieser Studie hat das Schreiben von mehr als 40 Wörtern (im Vgl. zu weniger als 40) den größten Effekt auf die Wahrscheinlichkeit, dass ein Lösungsvorschlag hinsichtlich der Glaubwürdigkeit positiv bewertet wird. Anschließend

haben die Faktoren der fehlenden Diskussionsbeiträge (Studie) und die Erfahrung und Bildung des Autors den größten Einfluss.

#### 5.3.5.2.4 Ordinale Regressionsanalyse für die Bewertungskategorie für den Grad der Klarheit der Beschreibung (K8)

Für die Variable, inwieweit der Lösungsvorschlag klar beschrieben ist, wurde ein gültiges ordinales Regressionsmodell für die 3er-Likert-Skala berechnet (Tabelle 9-7 in Anhang C). Das Regressionsmodell zeigt (Tabelle 9-15 in Anhang 0), dass die Prädiktoren für die Erfahrung des Bewertenden, die Anzahl der Wörter, die Studie, das Alter und die Erfahrung des Autors signifikant sind. In Tabelle 5-38 sind die Wahrscheinlichkeiten für die Ausprägungen der signifikanten unabhängigen Variablen abgebildet.

**Tabelle 5-38: Ordinales Regressionsmodell für das Kriterium den Grad der Klarheit der Beschreibung (K8). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird.**

<b>Bewertender: Erfahrung &gt; 1 Jahr</b>				
Bewertender: Erfahrung > 1 Jahr		negativ	neutral	positiv
<b>&lt; 1 Jahr</b> (n=35)	Mean	24,32%	24,53%	51,15%
	Std. Deviation	0,18506	0,08941	0,24864
<b>&gt; 1 Jahr</b> (n=291)	Mean	11,92%	16,35%	71,73%
	Std. Deviation	0,14903	0,09901	0,22367
<b>Anzahl der Wörter</b>				
Anzahl der Wörter		negativ	neutral	positiv
<b>0-40</b> (n=9)	Mean	50,03%	22,31%	27,66%
	Std. Deviation	0,32299	0,10741	0,22694
<b>41-80</b> (n=116)	Mean	10,58%	16,06%	73,36%
	Std. Deviation	0,12087	0,09881	0,20952
<b>81-120</b> (n=107)	Mean	15,34%	21,27%	63,39%
	Std. Deviation	0,13464	0,08696	0,20692
<b>121+</b> (n=94)	Mean	10,66%	13,58%	75,76%
	Std. Deviation	0,15516	0,10201	0,24134
<b>Studie</b>				
Studie		negativ	neutral	positiv
<b>Teilstudie 2.1</b> (n=120)	Mean	9,04%	13,31%	77,65%
	Std. Deviation	0,12234	0,10910	0,22167
<b>Teilstudie 2.2</b> (n=206)	Mean	15,71%	19,51%	64,79%
	Std. Deviation	0,17061	0,08880	0,22986
<b>Autor: Erfahrung &gt; 1 Jahr</b>				
Autor: Erfahrung > 1 Jahr		negativ	neutral	positiv
<b>&lt; 1 Jahr Erfahrung</b> (n=76)	Mean	27,74%	24,03%	48,23%
	Std. Deviation	0,22377	0,07809	0,25504
<b>&gt; 1 Jahr Erfahrung</b> (n=250)	Mean	8,85%	15,16%	75,99%
	Std. Deviation	0,09507	0,09832	0,18543



		Autor: Alter		
Autor: Alter		negativ	neutral	positiv
<b>Kein Alter angegeben</b> (n=10)	Mean	8,63%	18,54%	72,83%
	Std. Deviation	0,02190	0,03491	0,05676
<b>18-24</b> (n=33)	Mean	3,83%	8,99%	87,19%
	Std. Deviation	0,03485	0,06858	0,10319
<b>25-34</b> (n=116)	Mean	15,92%	15,84%	68,24%
	Std. Deviation	0,20877	0,11203	0,28598
<b>35-44</b> (n=67)	Mean	12,69%	18,04%	69,27%
	Std. Deviation	0,13746	0,09368	0,21557
<b>45-54</b> (n=59)	Mean	14,72%	21,78%	63,50%
	Std. Deviation	0,11820	0,07964	0,18735
<b>55-64</b> (n=41)	Mean	13,22%	19,58%	67,20%
	Std. Deviation	0,11838	0,09676	0,20967

Hinsichtlich der Klarheit eines Lösungsvorschlags wird dieses Kriterium häufig als positiv bewertet, wenn der Bewertende mehr als ein Jahr Erfahrung hat. Ein Grund dafür kann sein, dass dieser dann leichter den Lösungsvorschlag versteht und einordnen kann. Wenn der Beitrag mehr als 40 Wörter enthält, ist die Wahrscheinlichkeit für eine positive Bewertung hinsichtlich einer klaren Beschreibung des Lösungsvorschlags deutlich höher: ein Anstieg von ca. 45 %. Beim Autor sind die demografischen Daten Alter und Erfahrung signifikante Prädiktoren. Ein Beitrag wird häufiger als klar beschrieben bewertet, wenn der Autor mehr als 1 Jahr Erfahrung hat. Dort ist ein Anstieg von 28 % vorhanden! Hinsichtlich des Alters werden insbesondere die Lösungsvorschläge der Altersgruppe 18-24 als besonders klar beschrieben bewertet. Die übrigen Altersgruppen sind ungefähr ähnlich bewertet. Die größten Faktoren in diesem Modell sind die Anzahl der Wörter (sofern mehr als 40 Wörter geschrieben werden) sowie die Erfahrung des Autors und des Bewertenden.

#### 5.3.5.2.5 Ordinale Regressionsanalyse für die Bewertungskategorie für den Grad wie leicht der Lösungsvorschlag in die Praxis überführbar ist (K9)

Für die Variable wie leicht der Lösungsvorschlag in die Praxis überführbar ist, wurde ein gültiges ordinales Regressionsmodell für die 3er-Likert-Skala berechnet (Tabelle 9-7 in Anhang C). Im Regressionsmodell sind (Tabelle 9-16 in Anhang 0) die Prädiktoren für die Erfahrung und die Wochenarbeitsstunden des Bewertenden, die Anzahl der Wörter, die Studie, das Alter und die Erfahrung des Autors signifikant. In Tabelle 5-39 sind die Wahrscheinlichkeiten für die Ausprägungen der signifikanten unabhängigen Variablen abgebildet.

**Tabelle 5-39: Ordinales Regressionsmodell für das Kriterium den Grad wie leicht der Lösungsvorschlag in die Praxis überführbar ist (K9). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird.**

<b>Bewertender: Erfahrung &gt; 1 Jahr</b>				
Bewertender: Erfahrung > 1 Jahr		negativ	neutral	positiv
<b>&lt; 1 Jahr</b> (n=35)	Mean	19,22%	16,71%	64,07%
	Std. Deviation	0,16353	0,06071	0,20766
<b>&gt; 1 Jahr</b> (n=291)	Mean	13,96%	13,16%	72,88%
	Std. Deviation	0,15266	0,07650	0,20684
<b>Bewertender: Wochenarbeitsstunden</b>				
Bewertender: Wochenarbeitsstunden		negativ	neutral	positiv
<b>&lt; 1 Stunde / Woche</b> (n=42)	Mean	14,95%	13,42%	71,63%
	Std. Deviation	0,16824	0,07864	0,21792
<b>1 - 5 Stunden / Woche</b> (n=140)	Mean	18,59%	16,13%	65,28%
	Std. Deviation	0,16261	0,07165	0,21467
<b>5 - 15 Stunden / Woche</b> (n=65)	Mean	8,18%	9,21%	82,60%
	Std. Deviation	0,12112	0,06363	0,15732
<b>&gt; 15 Stunden / Woche</b> (n=79)	Mean	12,31%	12,59%	75,10%
	Std. Deviation	0,13624	0,07273	18,74%
<b>Anzahl der Wörter</b>				
Anzahl der Wörter		negativ	neutral	positiv
<b>0-40</b> (n=9)	Mean	53,63%	13,63%	32,74%
	Std. Deviation	0,38687	0,09654	0,29614
<b>41-80</b> (n=116)	Mean	11,58%	11,92%	76,49%
	Std. Deviation	0,12590	0,07533	0,19370
<b>81-120</b> (n=107)	Mean	16,30%	16,47%	67,23%
	Std. Deviation	0,11397	0,07140	0,17941
<b>121+</b> (n=94)	Mean	12,39%	12,20%	75,40%
	Std. Deviation	0,13929	0,06999	0,19984
<b>Studie</b>				
Studie		negativ	neutral	positiv
<b>Teilstudie 2.1</b> (n=120)	Mean	8,82%	10,22%	80,96%
	Std. Deviation	0,09897	0,06907	0,16307
<b>Teilstudie 2.2</b> (n=206)	Mean	17,85%	15,48%	66,68%
	Std. Deviation	0,17066	0,07275	0,21418
<b>Autor: Erfahrung &gt; 1 Jahr</b>				
Autor: Erfahrung > 1 Jahr		negativ	neutral	positiv
<b>&lt; 1 Jahr Erfahrung</b> (n=76)	Mean	28,17%	17,74%	54,09%
	Std. Deviation	0,23201	0,07103	0,25328
<b>&gt; 1 Jahr Erfahrung</b> (n=250)	Mean	10,37%	12,27%	77,36%
	Std. Deviation	0,08668	0,07250	0,15712

		Autor: Alter		
Autor: Alter		negativ	neutral	positiv
<b>Kein Alter angegeben</b> (n=10)	Mean	7,46%	10,48%	82,06%
	Std. Deviation	0,04670	0,05072	0,09726
<b>18-24</b> (n=33)	Mean	6,92%	8,99%	84,09%
	Std. Deviation	0,06890	0,07123	0,13906
<b>25-34</b> (n=116)	Mean	17,50%	12,89%	69,61%
	Std. Deviation	0,21175	0,07919	0,25578
<b>35-44</b> (n=67)	Mean	14,44%	14,33%	71,23%
	Std. Deviation	0,12860	0,07907	0,20021
<b>45-54</b> (n=59)	Mean	16,84%	17,06%	66,10%
	Std. Deviation	0,11083	0,06608	0,17144
<b>55-64</b> (n=41)	Mean	10,76%	13,45%	75,79%
	Std. Deviation	0,06885	0,05685	0,12446

In diesem Regressionsmodell ist die Erfahrung des Bewertenden ein signifikanter Prädiktor. Die Wahrscheinlichkeit auf einen als leicht in die Praxis überführbar bewerteten Lösungsvorschlag steigt um 8 %, wenn der Bewertende mehr als 1 Jahr Erfahrung hat (im Vgl. zu weniger als 1 Jahr Erfahrung). Hier kann ebenfalls die wachsende Erfahrung dazu beitragen, dass derjenige eher weiß, wie die Idee umzusetzen ist und sie daher als positiver bewertet. Das Weglassen einer Diskussion (Studie) erhöht die Wahrscheinlichkeit auf eine positive Bewertung hinsichtlich der Überführbarkeit in die Praxis um 14 %. Ein großer Sprung in der Wahrscheinlichkeit für positive Bewertung ist das Schreiben von mehr als 40 Wörtern in einem Lösungsvorschlag. Beim Autor haben die demografischen Werte hinsichtlich Alter und Erfahrung einen signifikanten Einfluss. Mehr als 1 Jahr Erfahrung als Autor lässt die Wahrscheinlichkeit auf Lösungsvorschläge, die leicht in die Praxis überführbar sind um 23 % steigen (im Vgl. zu weniger als 1 Jahr Erfahrung). Beim Alter haben die 18-24 ebenfalls eine höhere Wahrscheinlichkeit als die anderen Altersgruppen. Aufgrund der niedrigen Zahl der Personen, die ihr Alter angegeben haben, wird der Wert hier nicht diskutiert. Die größten Einflüsse auf einen als leicht in die Praxis umzusetzenden Lösungsvorschlag haben Anzahl der Wörter (sofern es mehr als 40 Wörter sind), sowie die Erfahrung des Autors und des Bewertenden.

#### 5.3.5.2.6 Ordinale Regressionsanalyse für die Bewertungskategorie für den Grad wie sehr der Lösungsvorschlag Einfluss auf die künftigen Handlungen des Bewertenden hat (K10)

Für die Variable inwieweit der Lösungsvorschlag Einfluss auf die künftigen Handlungen des Bewertenden hat, wurde ein gültiges ordinales Regressionsmodell für die 3er-Likert-Skala berechnet (Tabelle 9-7 in Anhang C). Das Regressionsmodell zeigt (Tabelle 9-17 in Anhang 0), dass die Prädiktoren für das Alter und die Bildung des Bewertenden, die Anzahl der Wörter und die Studie signifikant sind. In Tabelle 5-40 sind die

Wahrscheinlichkeiten für die Ausprägungen der signifikanten unabhängigen Variablen abgebildet.

**Tabelle 5-40: Ordinales Regressionsmodell für das Kriterium den Grad wie sehr der Lösungsvorschlag eine Auswirkung auf die künftigen Handlungen des Bewertenden hat (K10).** Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird.

<b>Bewertender: Alter</b>				
Bewertender: Alter		negativ	neutral	positiv
<b>17-24</b>	Mean	21,91%	25,04%	53,05%
(n=23)	Std. Deviation	0,08863	0,04494	0,13048
<b>25-34</b>	Mean	32,52%	25,62%	41,86%
(n=86)	Std. Deviation	0,17207	0,04615	0,17319
<b>35-44</b>	Mean	30,65%	24,83%	44,53%
(n=76)	Std. Deviation	0,18784	0,05550	0,18115
<b>45-54</b>	Mean	29,81%	24,88%	45,31%
(n=71)	Std. Deviation	0,17598	0,05030	0,18253
<b>55+</b>	Mean	14,78%	19,37%	65,84%
(n=70)	Std. Deviation	0,11736	0,05265	0,14568
<b>Bewertender: Akademiker (Bildung)</b>				
Bewertender: Akademiker (Bildung)		negativ	neutral	positiv
<b>No Academics</b>	Mean	29,74%	24,31%	45,95%
(n=173)	Std. Deviation	0,18569	0,05771	0,19524
<b>Academics</b>	Mean	23,77%	23,42%	52,81%
(n=153)	Std. Deviation	0,15722	0,05324	0,18053
<b>Anzahl der Wörter</b>				
Anzahl der Wörter		negativ	neutral	positiv
<b>0-40</b>	Mean	89,05%	7,40%	3,55%
(n=9)	Std. Deviation	0,04112	0,02666	0,01447
<b>41-80</b>	Mean	23,73%	23,96%	52,32%
(n=116)	Std. Deviation	0,13891	0,04900	0,17413
<b>81-120</b>	Mean	30,70%	25,73%	43,56%
(n=107)	Std. Deviation	0,15174	0,03979	0,17367
<b>121+</b>	Mean	20,66%	23,30%	56,04%
(n=94)	Std. Deviation	0,11317	0,05449	0,16193
<b>Studie</b>				
Studie		negativ	neutral	positiv
<b>Teilstudie 2.1</b>	Mean	22,01%	22,38%	55,61%
(n=120)	Std. Deviation	0,16845	0,05688	0,17891
<b>Teilstudie 2.2</b>	Mean	29,80%	24,78%	45,42%
(n=206)	Std. Deviation	0,17307	0,05326	0,18865

Hinsichtlich des Kriteriums, inwieweit der Lösungsvorschlag eine Auswirkung auf die künftige Handlung des Bewertenden hat, sind interessanterweise nur demografische Variablen des Bewertenden und nicht der Autoren unter den signifikanten Prädiktoren. Beim Alter der Bewertenden ist die Wahrscheinlichkeit auf eine positive Bewertung in

der jungen (17-24) und in der älteren (55+) Altersgruppe am höchsten. Die Wahrscheinlichkeit ist bei den übrigen Altersgruppen ungefähr ähnlich. Akademiker bewerten die Lösungsvorschläge ebenfalls dahingehend als wahrscheinlicher, dass sie einen Einfluss auf die eigenen Handlungen haben. Der größte Anstieg der Wahrscheinlichkeit auf eine positive Bewertung in diesem Regressionsmodell liegt bei dem Schreiben von mehr als 40 Wörtern in einem Lösungsvorschlag. Das Fehlen von anderen Diskussionsbeiträgen bei der Entwicklung eines Lösungsvorschlags erhöht die Wahrscheinlichkeit auf eine positive Bewertung hinsichtlich der Auswirkung auf künftige Handlungen des Bewertenden um 10 %.

#### 5.3.5.2.7 Ordinale Regressionsanalyse für die Bewertungskategorie für den Grad wie sehr der Bewertende den Lösungsvorschlag weiterempfehlen würde (K11)

Für die Variable für den Grad, wie sehr der Bewertende den Lösungsvorschlag weiterempfehlen würde, wurde ein gültiges ordinale Regressionsmodell für die 3er-Likert-Skala berechnet (Tabelle 9-7 in Anhang C). Im Regressionsmodell sind (Tabelle 9-18 in Anhang 0) die Prädiktoren für das Alter des Bewertenden, die Anzahl der Wörter, die Studie, das Alter und die Erfahrung des Autors signifikant. In Tabelle 5-41 sind die Wahrscheinlichkeiten für die Ausprägungen der signifikanten unabhängigen Variablen abgebildet.

Das Fehlen von Diskussionsbeiträgen beim Verfassen eines Lösungsvorschlags erhöht die Chance auf eine Weiterempfehlung des Lösungsvorschlags seitens des Bewertenden um ca. 14 %. Den größten Einfluss hat die Erfahrung des Autors, welche die Wahrscheinlichkeit um 20 % erhöht sobald mehr als 1 Jahr Erfahrung vorliegt. Dies kann implizieren, dass dann der Lösungsvorschlag für den Bewertenden als besonders wertvoll angesehen wird. Bei der Anzahl der Wörter steigt die Wahrscheinlichkeit ebenfalls stark, sobald mehr als 40 Wörter geschrieben werden. Sowohl das Alter des Autors als auch das Alter des Bewertenden sind signifikante Prädiktoren. Bei beiden zeichnet sich ein ähnliches Bild, bei dem die Wahrscheinlichkeit auf eine Weiterempfehlung steigt, wenn der Autor entweder jünger (17-24) oder älter (55+) ist.

**Tabelle 5-41: Ordinales Regressionsmodell für das Kriterium den Grad wie sehr der Bewertende den Lösungsvorschlag weiterempfehlen würde (K11). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird.**

<b>Bewertender: Alter</b>				
Bewertender: Alter		negativ	neutral	positiv
<b>17-24</b>	Mean	16,94%	12,76%	70,30%
(n=23)	Std. Deviation	0,14475	0,04810	0,18105
<b>25-34</b>	Mean	25,55%	14,81%	59,65%
(n=86)	Std. Deviation	0,20215	0,05341	0,22003
<b>35-44</b>	Mean	36,81%	16,51%	46,68%
(n=76)	Std. Deviation	0,23041	0,04661	0,22257
<b>45-54</b>	Mean	27,98%	15,37%	56,65%
(n=71)	Std. Deviation	0,20632	0,04764	0,22438
<b>55+</b>	Mean	16,93%	12,40%	70,67%
(n=70)	Std. Deviation	0,15848	0,05431	0,18479
<b>Anzahl der Wörter</b>				
Anzahl der Wörter		negativ	neutral	positiv
<b>0-40</b>	Mean	86,86%	6,54%	6,59%
(n=9)	Std. Deviation	0,13502	0,06209	0,07362
<b>41-80</b>	Mean	23,69%	13,26%	63,06%
(n=116)	Std. Deviation	0,22150	0,05329	0,23889
<b>81-120</b>	Mean	26,46%	16,09%	57,45%
(n=107)	Std. Deviation	0,16769	0,04755	0,20114
<b>121+</b>	Mean	23,35%	15,56%	61,09%
(n=94)	Std. Deviation	0,14954	0,04442	0,18700
<b>Studie</b>				
Studie		negativ	neutral	positiv
<b>Teilstudie 2.1</b>	Mean	18,66%	13,23%	68,11%
(n=120)	Std. Deviation	0,16031	0,04831	0,18804
<b>Teilstudie 2.2</b>	Mean	30,66%	15,50%	53,83%
(n=206)	Std. Deviation	0,22262	0,05298	0,23292
<b>Autor: Erfahrung &gt; 1 Jahr</b>				
Autor: Erfahrung > 1 Jahr		negativ	neutral	positiv
<b>&lt; 1 Jahr Erfahrung</b>	Mean	40,98%	15,49%	43,52%
(n=76)	Std. Deviation	0,26796	0,05639	0,24559
<b>&gt; 1 Jahr Erfahrung</b>	Mean	21,76%	14,41%	63,82%
(n=250)	Std. Deviation	0,16492	0,05098	0,19991
<b>Autor: Alter</b>				
Autor: Alter		negativ	neutral	positiv
<b>Kein Alter angegeben</b>	Mean	8,44%	9,06%	82,51%
(n=10)	Std. Deviation	0,03881	0,03447	0,07318
<b>18-24</b>	Mean	15,77%	12,62%	71,61%
(n=33)	Std. Deviation	0,11703	0,06059	0,17371
<b>25-34</b>	Mean	33,45%	14,44%	52,11%
(n=116)	Std. Deviation	0,26681	0,05374	0,26004

<b>35-44</b>	Mean	24,50%	15,08%	60,42%
(n=67)	Std. Deviation	0,17785	0,05027	0,21136
<b>45-54</b>	Mean	29,80%	17,30%	52,90%
(n=59)	Std. Deviation	0,15912	0,04228	0,19108
<b>55-64</b>	Mean	16,37%	13,86%	69,77%
(n=41)	Std. Deviation	0,09224	0,04213	0,13067

#### 5.3.5.2.8 Ordinale Regressionsanalyse für die Bewertungskategorie inwieweit der Bewertende den Lösungsvorschlag nutzen würde, wenn er das Problem hätte (K12)

Für die Variable inwieweit der Bewertende den Lösungsvorschlag nutzen würde, wurde ein gültiges ordinales Regressionsmodell für die 3er-Likert-Skala berechnet (Tabelle 9-7 in Anhang C). Das Regressionsmodell zeigt (Tabelle 9-19 in Anhang 0), dass die Prädiktoren für die Anzahl der Wörter, die Studie und die Erfahrung des Autors signifikant sind. In Tabelle 5-42 sind die Wahrscheinlichkeiten für die Ausprägungen der signifikanten unabhängigen Variablen abgebildet.

**Tabelle 5-42: Ordinales Regressionsmodell für das Kriterium den Grad wie sehr der Bewertende den Lösungsvorschlag nutzen würde, sofern er das Problem hätte (K12). Jede Zeile zeigt die Wahrscheinlichkeit für die jeweilige Ausprägung der unabhängigen Variable an, mit der ein Lösungsvorschlag negativ, neutral oder positiv bewertet wird.**

		<b>Anzahl der Wörter</b>		
Anzahl der Wörter		negativ	neutral	positiv
<b>0-40</b>	Mean	75,91%	14,24%	9,85%
(n=9)	Std. Deviation	0,19376	0,10483	0,09080
<b>41-80</b>	Mean	20,90%	21,14%	57,96%
(n=116)	Std. Deviation	0,17043	0,06102	0,20327
<b>81-120</b>	Mean	27,71%	25,51%	46,78%
(n=107)	Std. Deviation	0,13805	0,05410	0,17617
<b>121+</b>	Mean	19,37%	22,76%	57,87%
(n=94)	Std. Deviation	0,10958	0,05758	0,16157
		<b>Studie</b>		
Studie		negativ	neutral	positiv
<b>Teilstudie 2.1</b>	Mean	17,92%	20,49%	61,59%
(n=120)	Std. Deviation	0,13975	0,06087	0,18576
<b>Teilstudie 2.2</b>	Mean	27,88%	24,23%	47,90%
(n=206)	Std. Deviation	0,17924	0,06106	0,19259
		<b>Autor: Erfahrung &gt; 1 Jahr</b>		
Autor: Erfahrung > 1 Jahr		negativ	neutral	positiv
<b>&lt; 1 Jahr Erfahrung</b>	Mean	37,05%	24,32%	38,63%
(n=76)	Std. Deviation	0,22165	0,06558	0,19979
<b>&gt; 1 Jahr Erfahrung</b>	Mean	20,31%	22,41%	57,29%
(n=250)	Std. Deviation	0,13207	0,06232	0,18047

Das Fehlen von Diskussionsbeiträgen (*Studie*) erhöht die Wahrscheinlichkeit auf einen Lösungsvorschlag, den der Bewertende selbst nutzen würde, wenn er das Problem hätte (Anstieg um 13 %). Die Anzahl der Wörter haben hier ebenfalls einen Effekt auf die Wahrscheinlichkeit von positiv bewerteten Beiträgen, sofern mehr als 40 Wörter geschrieben werden. Die einzige demografische Variable, die in diesem Modell ein signifikanter Prädiktor ist, ist die Erfahrung des Autors. Wenn der Autor erfahren ist (mehr als 1 Jahr Erfahrung), dann steigt die Wahrscheinlichkeit auf einen Lösungsvorschlag, den der Bewertende selbst nutzen würde, um ca. 19 %. Somit hat in diesem Modell die Erfahrung des Autors die größte Auswirkung auf eine positive Bewertung.

#### 5.3.5.2.9 Übersicht der Einflüsse der demografischen Daten

Es muss hervorgehoben werden, dass in der Ausprägung von 0-40 Wörtern in der unabhängigen Variable für die Anzahl der Wörter nur 9 Beiträge enthalten sind. Es lässt sich vermuten, dass diese sehr niedrige Anzahl der Grund für den statistischen Effekt ist. Dadurch sollte dieser Fund vorsichtig bewertet werden. Dennoch fällt auf, dass in allen Regressionsmodellen die Wahrscheinlichkeit auf eine positive Bewertung für die jeweilige Bewertungskategorie für die Ausprägung 41-80 Wörter und 121+ Wörter ähnlich hoch ist. Für die Ausprägung 81-120 Wörter ist diese Wahrscheinlichkeit niedriger, wenn auch mit relativ geringem Unterschied. Dieses Ergebnis scheint zu implizieren, dass eine Mindestlänge eines Lösungsvorschlags vorliegen muss, um eine höhere Wahrscheinlichkeit auf positive Bewertungen zu bekommen und dass längere Beiträge ebenfalls eine höhere Wahrscheinlichkeit haben. Dies kann bspw. daran liegen, dass in den längeren Beiträgen mehr Informationen enthalten sind, was zu einer positiveren Bewertung führt.

Beim Alter des Autors ist in dieser Betrachtung (Regressionsmodelle für die Bewertungskategorien unter Berücksichtigung der Lösungs-Prompts) keine Person in der Altersgruppe von 65+ enthalten. Aus diesem Grund fehlt diese Kategorie in den obigen Tabellen.

In fast allen Kategorien mit einem Regressionsmodell erhöht das Fehlen von Diskussionsbeiträgen (Variable *Studie*) die Wahrscheinlichkeit auf eine höhere Bewertung. Dieser Effekt lag häufig zwischen 10 und 15 % und ist als eher gering zu bewerten.

Ebenfalls ein wichtiger Faktor für eine positive Bewertung in den Bewertungskategorien ist die Erfahrung des Autors. Hier erhöht eine größere Erfahrung (>1 Jahr) im Vergleich zu wenig Erfahrung (< 1 Jahr) die Wahrscheinlichkeit auf eine höhere Bewertung



zwischen 7 und 28 %. Damit ist der Effekt größer als die Auswirkung der Diskussionsbeiträge. Dieses Ergebnis ist nachvollziehbar, da Crowdworker zu einem geläufigen und ihnen bekannten Problem einen Lösungsvorschlag entwickeln sollten. Offenbar ist das gewählte Problem in der Herausforderung nicht trivial und erfahrenere Crowdworker können hier eher zu qualitativen Lösungsvorschlägen beitragen.

Das Alter des Bewertenden und des Autors waren häufiger ein signifikanter Prädiktor in den Regressionsmodellen. Hier zeigte sich in der Regel, dass insbesondere bei den Autoren die jüngeren (18-24) und die älteren (55-64) Autoren Lösungsvorschläge mit einer höheren Bewertung schreiben bzw. erfassen.

Mehrere demografische Faktoren haben keinerlei Einflüsse bei den Regressionsmodellen. Dazu gehören das Geschlecht und das Land des Bewertenden. Beim Autor ist lediglich das Geschlecht nie ein signifikanter Prädiktor gewesen. Demografische Merkmale wie das Land bzw. die Wochenarbeitsstunden des Autors war nur im Regressionsmodell für das Kriterium hinsichtlich ungewöhnlicher Elemente enthalten (K2).

Die häufigsten Prädiktoren in den Regressionsmodellen sind die Anzahl der Wörter und die Erfahrung und das Alter des Autors, sowie das Vorhandensein von Diskussionsbeiträgen (Studie). Somit zeigt sich, dass hier das Vorhandensein bzw. Nichtvorhandensein von Diskussionsbeiträgen nicht der alleinige Grund für die in Abschnitt 5.3.5.1 ermittelten Unterschiede hinsichtlich der Bewertung zwischen den beiden Teilstudien sind. Insbesondere die Erfahrung des Autors hatte häufig eine größere Wirkung auf die positive Bewertung eines Beitrags.

### **5.3.5.3 Zusammenhänge zwischen Bewertungskriterien und reflexionsbezogenen Codes**

In diesem Abschnitt wird untersucht, inwieweit die reflexionsbezogenen Codes einen Zusammenhang zu den Bewertungskategorien und den einzelnen Bewertungen aufweisen. Tabelle 5-43 zeigt die Ergebnisse eines  $\chi^2$ -Tests. In dieser Teilstudie wurden erfahrungsbasierte Lösungsvorschläge bewertet. Dementsprechend hat jeder Beitrag in dieser Tabelle zusätzlich den Code SUG\_EXP enthalten. Eine Kalkulation für den Code SUG\_EXP wurde nicht durchgeführt, da dieser konstant ist. Da der Code SUG\_EXP, für einen erfahrungsbasierten Lösungsvorschlag, immer in Zusammenhang mit dem Code für Erfahrung (EXP) vergeben wird (siehe Abschnitt 3.4.1), ist das Auftauchen des Codes EXP in dieser Tabelle vorhersehbar.

Beiträge, in denen zusätzlich ein Code für Erfahrung (EXP), Wissen (KNO) und wissensbasierte Lösungsvorschläge (SUG\_KNO) enthalten sind, werden überwiegend als glaubwürdig bewertet (K7). Hinsichtlich der Codes für Wissen und wissensbasierte Lösungsvorschläge sollte dieses Ergebnis vorsichtig bewertet werden, da bei der Datenaufbereitung die Lösungsvorschläge dahingehend bereinigt wurden. Am häufigsten ist jedoch eine normale Zustimmung (*agree*) und keine starke Zustimmung (*strongly agree*). Hier werden die Ergebnisse hinsichtlich der Glaubwürdigkeit bewertet positiver, wobei Raum für Verbesserung in Form von besseren Bewertungen besteht.

Die Tabelle 5-43 zeigt auch, dass ein Beitrag, der zusätzlich Erfahrungen enthält, eher danach bewertet wird, dass er weniger Elemente enthält, an die man selbst nicht gedacht hat (K3). Dementsprechend lässt sich implizieren, dass diese Beiträge wenig Neues für den Empfänger bereithalten und diese die Erfahrungen ggf. selbst erlebt haben oder davon gehört haben. In diesem Setting scheinen den Bewertenden demnach die meisten Vorschläge bekannt gewesen zu sein. Wenn ein Beitrag zusätzlich den Code für Wissen (KNO) enthält, ist die Neuheit relativ gleichmäßig verteilt, wobei etwas mehr Personen die Beiträge als nicht neu bewerten (K1). Beiträge die zusätzlich den Code für Erfahrungen (EXP) beinhalten werden seltener als ungewöhnlich bewertet (K2). Dies impliziert, dass die Ideen an dieser Stelle überwiegend als üblich verstanden werden. Gleichzeitig werden Beiträge, in denen der Code für Hinweise (ADV) zusätzlich enthalten ist, so bewertet.

**Tabelle 5-43: Verteilungen der Bewertungen innerhalb der einzelnen reflexionsbezogenen Codes aufgeteilt nach Bewertungskategorie (5er-Likert-Skala).**

Rating	Code	$\chi^2$ (df=4)	Strongly disagree	disagree	Neither nor	agree	Strongly agree
<b>Suggestion is Believable (K7)</b>	EXP	$\chi^2=9,311$ , $p=,054$	2,3 %	6,5 %	12 %	53 %	25,7 %
<b>Suggestion is Believable (K7)</b>	KNO	$\chi^2=14,811$ $p=,005$	2,8 %	7,1 %	12,2 %	47,1 %	30,9 %
<b>Suggestion is Believable (K7)</b>	SUG_KNO	$\chi^2=9,704$ , $p=,045$	4,9 %	5,6 %	13,6 %	46,9 %	29 %
<b>Elements I didn't think of (K3)</b>	EXP	$\chi^2=11,307$ , $p=,023$	19,1 %	47,7 %	14,4 %	15,6 %	3,2 %
<b>Novel (K1)</b>	KNO	$\chi^2=10,384$ , $p=,034$	10,9 %	30,1 %	21,3 %	29,6 %	8,1 %
<b>Unusual (K2)</b>	EXP	$\chi^2=15,590$ , $p=,004$	23,8 %	44,6 %	15,6 %	13,3 %	2,7 %
<b>Unusual (K2)</b>	ADV	$\chi^2=11,789$ , $p=,019$	28,8 %	49,4 %	11,8 %	8,2 %	1,8 %

### 5.3.5.4 Analyse der Prompts

Es soll verglichen werden, ob die bei den unterschiedlichen Prompts entstandenen erfahrungsbasierten Lösungsvorschläge in den Teilstudien 2.1 und 2.2 unterschiedliche Qualitäten aufweisen. Tabelle 5-44 zeigt, dass der Lösungs-Prompt im Vergleich zum Erfahrungs-Prompt qualitativ hochwertigere Lösungen hervorbringt. Diese Lösungen passen eher auf das Problem (K4), lösen laut dem Bewertenden eher das Problem (K5), sind hilfreicher (K6), und eher in die Praxis zu überführen (K9) bzw. haben eher einen Einfluss auf die künftigen Handlungen des Bewertenden (K10). Zusätzlich werden sie häufiger weiterempfohlen (K11). Es gibt demnach keine Unterschiede hinsichtlich des Grads der Neuheit (K1), inwieweit die Lösung ungewohnt ist (K2), ob sie Komponenten enthält, die der Bewertende selbst nicht gedacht hat (K3), oder der Klarheit der Formulierung (K8). Ebenfalls unterscheiden sich die Lösungen nicht hinsichtlich der Glaubwürdigkeit (K7).

**Tabelle 5-44: Mann-Whitney-U-Test zum Vergleich der Qualitäten der erfahrungsbasierten Lösungsvorschläge aus Teilstudie 2.1 generiert aus Lösungs-Prompts und Erfahrungs-Prompts (n=300)**

	5er-Likert-Skala		3er-Likert-Skala		Prompts
	U	Sig.	U	Sig.	
Fits the issue (K4)	9109	,013	8955	,001	LP > EP
Likely solves the issue (K5)	8644,5	,002	8705,5	,002	LP > EP
Helpful (K6)	8764	,003	9085	,006	LP > EP
Put into practice (K9)	9080	,012	9491	,021	LP > EP
Impact (K10)	9394,5	,046	9621,5	,080	LP > EP
Recommend (K11)	9060	,013	9165,5	,011	LP > EP

Interessanterweise zeigen Mann-Whitney-U-Tests bei einem Vergleich zwischen dem Lösungs-Prompt (LP) und dem Kontroll-Prompt (KP) bzw. dem Erfahrungs-Prompt (EP) und dem Kontroll-Prompt keine signifikanten Unterschiede in den Bewertungskriterien. Daher scheinen die Bewertungen des Kontroll-Prompts zwischen denen der Lösungs- bzw. Erfahrungs-Prompts zu liegen. Die Bewertung der Kontroll-Prompts scheint aber leicht höher zu sein, da zwischen dem Lösungs-Prompt und dem Kontroll-Prompt keine signifikanten Zusammenhänge gefunden wurden, aber zwischen dem Lösungs-Prompt und dem Erfahrungs-Prompt solche signifikanten Zusammenhänge gefunden wurden. Um einzuschätzen, wie viel besser die Lösungsvorschläge des Lösungs-Prompts bewertet wurden, wurde der Mittelwert der Bewertungen samt Standardabweichung berechnet (siehe Tabelle 5-45). Die Tabelle zeigt, dass die Bewertungskategorien im Durchschnitt ca. 0,33 Einheiten niedriger sind und die Standardabweichung um ca. 0,15 größer ist, sprich die Werte leicht mehr streuen. Die Werte sind jedoch bei allen

Bewertungskategorien um den Wert drei und sehr oft etwas höher angesiedelt. Damit liegen die Ideen in einem Bereich, bei dem die Bewertenden weder positiv noch negativ geurteilt haben. Bei dem Lösungs-Prompt gibt es nur zwei Werte, die im Bereich von vier liegen, sprich der Bewertung agree.

**Tabelle 5-45: Vergleich der Bewertungen für den Lösungs- und Erfahrungs-Prompt in Teilstudie 2.1, bei denen sich die Bewertungen signifikant unterscheiden (5er-Likert-Skala)**

	Lösungs-Prompt		Erfahrungs-Prompt	
	Mittelwert	St. Abw.	Mittelwert	St. Abw.
Fits the issue (K4)	3,98	0,840	3,67	1,061
Likely solves the issue (K5)	3,35	0,993	2,96	1,133
Helpful (K6)	3,83	0,990	3,46	1,130
Put into practice (K9)	4,00	0,987	3,71	1,087
Impact (K10)	3,43	1,067	3,14	1,182
Recommend (K11)	3,65	1,142	3,31	1,216

Aus Teilstudie 2.2 sind nur die Ergebnisse von zwei Prompts eingeflossen (der Prompt, der expliziten Bezug auf Diskussionskommentare erfordert wurde ausgelassen; siehe Abschnitt 5.3.2). Hier wurde ebenfalls ein Vergleich durchgeführt, um zu prüfen, inwieweit sich die aus den Prompts resultierenden Beiträge qualitativ unterscheiden. Es wurden der Lösungs-Prompt (in Teilstudie 2.1 und 2.2 identisch) mit dem Kontroll-Lösungs-Prompt verglichen. Die Tabelle 5-46 zeigt, dass sich die Qualität der Beiträge bei den beiden Prompts nur in einem Aspekt unterscheiden: Dem Grad, wie sehr der Vorschlag auf das Problem passt (K4). Dies bedeutet, dass der Kontroll-Lösungs-Prompt (KLP) in diesem Studienaufbau qualitativ größtenteils gleichwertige Ergebnisse hervorbringt. Die zusätzliche Formulierung im Lösungs-Prompt hilft bei der Hinführung auf eine zu dem Problem passendere Formulierung.

**Tabelle 5-46: Mann-Whitney-U-Test zum Vergleich der Qualitäten der erfahrungsbasierten Lösungsvorschläge aus Teilstudie 2.2, berechnet aus Kontroll-Lösungs-Prompt und Lösungs-Prompt (n=384)**

	5er-Likert-Skala		3er-Likert-Skala		Prompts
	U	Sig.	U	Sig.	
Fits the issue (K4)	16112,5	,036	16084,5	,015	KLP > LP

### 5.3.5.5 Weitere Analysen

Bei der Analyse der Anzahl der Wörter im Kontext der Bewertung der Lösungsvorschläge lässt sich feststellen, dass die Anzahl der Wörter in fast allen Bewertungskategorien eine

Rolle spielen (bis auf Elemente, an die man nicht selbst gedacht hat (K3), und ob der Lösungsvorschlag ungewöhnlich ist (K2); siehe Anhang 0 in

Tabelle 9-10). In den zehn anderen Bewertungskategorien findet sich ein signifikanter Zusammenhang zwischen der Anzahl der Wörter und der Bewertung. Bei acht von zehn Kategorien (außer Neuheit (K1) und Weiterempfehlung (K11)) stieg bei jeder Kategorie mit mehr Wörtern der Anteil von positiv bewerteten Lösungsvorschlägen. Bei den zwei genannten Kategorien für Neuheit (K1) und Weiterempfehlung (K11) hatte die höchste Kategorie (121+ Wörter) leicht weniger Anteil an positiv bewerteten Lösungsvorschlägen als die vorherige Kategorie (81-120 Wörter). Daraus lässt sich ableiten, dass in acht von insgesamt zwölf Kategorien, ein Beitrag besser bewertet wird, je mehr Wörter er enthält. Somit trägt die Anzahl der Wörter dazu bei, dass der Lösungsvorschlag besser auf das Problem passt (K4), das Problem wahrscheinlicher löst (K5), hilfreich ist (K6), glaubwürdig ist (K7), klar beschrieben ist (K8), und der Bewertende meint den Lösungsvorschlag in die Praxis umsetzen zu können (K9). Zusätzlich vermuten die Bewertenden bei mehr Wörtern, dass der Lösungsvorschlag eher einen Einfluss auf ihr zukünftiges Verhalten hat (K10) und dass sie den Lösungsvorschlag eher weiterempfehlen würden (K11). Mögliche Ideen, wie diese Ergebnisse in der Praxis genutzt werden können, werden in Abschnitt 6.1.7 diskutiert.

Ein  $\chi^2$ -Test für die Lösungs-Prompts ergab keinen Unterschied in der Anzahl der Wörter für die beiden Konditionen mit und ohne Diskussionsbeiträge. Allerdings kann festgestellt werden, dass in Teilstudie 2.1 der Lösungs-Prompt durchschnittlich mehr Wörter als der Kontroll-Lösungs-Prompt hervorbringt ( $U=15267$ ,  $p=,003$ ). In Teilstudie 2.1 hingegen hat der Lösungs-Prompt durchschnittlich mehr Wörter als der Erfahrungs-Prompt hervorgerufen ( $U=8463$ ,  $p=,001$ ). In Teilstudie 2.1 gibt es keine weiteren Unterschiede.

Es wurde ebenfalls die Anzahl der Codes für den Lösungs-Prompt verglichen, der in den beiden Teilstudien 2.1 und 2.2 identisch war. Von allen reflexionsbezogenen Codes ist lediglich bei dem Code für Zustimmung (AGR) ein signifikanter Unterschied vorhanden ( $\chi^2=8,891$ ,  $p=,003$ ). Der Code wurde in Teilstudie 2.1 häufiger vergeben, in der Diskussionsbeiträge sichtbar waren. Ein Vergleich aller Prompts ist nicht sinnvoll, da bspw. das Vorhandensein eines Prompts für Erfahrungen in Teilstudie 2.1 zu mehr Erfahrungen führen würde.

### 5.3.6 Diskussion

In diesem Abschnitt werden zunächst die einzelnen Bewertungskriterien diskutiert. Anschließend erfolgt eine kritische Betrachtung der übrigen Ergebnisse. Es werden Hinweise zum Design von Unterstützungsfunktionen für Reflexion diskutiert. Es wird ebenfalls die Forschungsfrage der Teilstudie diskutiert, inwieweit sich die Qualität von Lösungsvorschlägen unterscheidet, wenn Lösungs-Prompts in Diskussionen mit und ohne Diskussionsbeiträgen genutzt werden.

#### 5.3.6.1 Überprüfung der Kriterien

Zunächst erfolgt eine sukzessive Diskussion der einzelnen Bewertungskriterien für Lösungsvorschläge.

Zwei Kategorien haben abgefragt, ob ein Lösungsvorschlag neu ist (K1) oder ob er ungewöhnliche Elemente enthält (K2). Es zeigt sich, dass es keine Unterschiede zwischen den Konditionen eines Themas mit oder ohne Diskussionsbeitrag gibt (siehe Tabelle 5-33). Der Fakt, dass für die Studie nur Crowdworker mit dem höchsten Rang an Erfahrung zugelassen wurden und die überwiegende Mehrheit der Probanden mehr als ein Jahr Erfahrung als Crowdworker hat (siehe Abbildung 5-13), lässt vermuten, dass die Vorschläge der erfahrenen Probanden (vgl. Abbildung 5-3 bzw. Abbildung 5-8) für die ebenso erfahrenen anderen Probanden weder viel Neues noch Überraschendes beinhalten. Ein Mann-Whitney-U-Test bestätigt diese Vermutung und zeigt keinen signifikanten Unterschied bei der Bewertung der Beiträge zwischen den Gruppen, wenn nur Personen (Autoren und Bewertende) mit mehr als einem Jahr Erfahrung betrachtet werden. Ebenso haben die Prompts diese Aspekte nicht explizit angesprochen. Diese Erfahrung ist vermutlich der Grund, warum es bei dem Kriterium, inwieweit der Lösungsvorschlag Elemente enthält, an die man selbst nicht gedacht hat (K3), ebenfalls keinen Unterschied zwischen den Konditionen gibt. Dementsprechend lässt sich festhalten, dass die Lösungsvorschläge in beiden Konditionen über ähnlich viel Neues und Überraschendes verfügen und Dinge enthalten, an die der Bewertende nicht gedacht hat. Dennoch ist die durchschnittliche Bewertung dieser Kriterien im Vergleich zu den anderen Kriterien am niedrigsten (siehe Tabelle 5-31). Dieser fehlende Unterschied bei den Bewertungen stärkt die Wirkung der Prompts, da weniger „Nebeneffekte“ auftauchen.

Die Tabelle 5-33 zeigt, dass bei den Kriterien K4 bis K12 jeweils die Lösungsvorschläge aus dem Thema ohne Diskussion (abgebildet in Teilstudie 2.1) durchschnittlich höher bewertet wurden als die Lösungsvorschläge in dem Thema mit Diskussionsbeiträgen (abgebildet in Teilstudie 2.2). Insgesamt zeigt sich aus der Analyse, dass der Prompt für

Lösungen in einem neuen Thema ohne Diskussionsbeiträge qualitativ hochwertigere Lösungsvorschläge hervorbringt als in einem Thema mit Diskussionsbeiträgen. Dies ist überraschend und passt nicht zur Literatur zu kollaborativer Reflexion, bei der eine Diskussion ein wichtiger Bestandteil ist, bevor Lösungen gemeinsam entwickelt werden (Dyke 2006; Nicholson u. a. 2003; Woerkom und Croon 2008; Kim und Lee 2002). Der Unterschied in der Bewertung ist nicht sonderlich hoch und aus diesem Grund können die Prompts in beiden Situationen in diesem Szenario verwendet werden. Weitere Studien sind notwendig, um zu evaluieren, inwieweit anders gestaltete Prompts ähnliche oder qualitative Lösungsvorschläge in einer Situation mit Lösungsvorschlägen hervorrufen können.

Offen ist, wie sich die Qualität der Lösungsvorschläge, die aus einem Lösungs-Prompt (LP) resultieren, von den Lösungsvorschlägen der beiden anderen Prompts unterscheiden. Aus Tabelle 5-44 lässt sich ableiten, dass der Lösungs-Prompt in Teilstudie 2.1 signifikant besser bewertete Lösungen hervorruft als der Erfahrungs-Prompt (EP), was die Kriterien zur Passung des Vorschlags auf das Problem (K4), wie wahrscheinlich es das Problem löst (K5), wie hilfreich der Lösungsvorschlag ist (K6), wie leicht er in die Praxis überführbar ist (K9), wie sehr der Lösungsvorschlag eine Auswirkung auf das eigene Verhalten hat (K10) und ob man den Lösungsvorschlag weiterempfehlen würde (K11), betrifft. Die gefundenen Unterschiede bestanden nur zwischen dem Lösungs-Prompt und dem Erfahrungs-Prompt. Zwischen den anderen Prompts (LP und KP, bzw. EP und KP) wurden keine signifikanten Unterschiede gefunden. Dies bedeutet, dass die Bewertungen wie folgt sortiert sind:  $LP > KP > EP$ .

Es folgt, dass der Lösungs-Prompt in diesem Szenario bessere Lösungen hervorbringt als der Erfahrungs-Prompt, aber nicht wesentlich schlechter ist als der Kontroll-Prompt. Es zeigt sich, dass sich der Lösungs-Prompt in einem Thema ohne Diskussionsbeiträge gut zur Anregung von Lösungsvorschlägen eignet.

#### **5.3.6.2 Diskussion der übrigen Ergebnisse**

Bei den Kriterien, inwieweit der Lösungsvorschlag glaubwürdig ist (K7) und ob man ihn anderen empfehlen würde (K11) finden sich signifikante Unterschiede in den Konditionen. Bei beiden Kriterien ist die Bewertung der Lösungsvorschläge, die aus dem Lösungs-Prompt in Themen ohne Diskussionsbeiträge entstanden sind, höher. Dies passt zu den übrigen Kriterien, aber nicht zur Literatur über Reflexion, bei der der eigenen Erfahrung immer ein hoher Stellenwert zugewiesen wird. Zumindest scheinen die Nutzer nicht von den vorhandenen Diskussionsbeiträgen profitiert zu haben.

Bei der Analyse der Bewertungen der Lösungsvorschläge hinsichtlich der Prompts zeigt Tabelle 5-44, dass die Lösungs-Prompts in Teilstudie 2.1 (ohne Diskussion) in mehreren Kategorien bessere Ergebnisse erzielen als die Erfahrungs-Prompts. Nachvollziehbar ist, dass wenn ein Prompt nach Lösungen fragt, diese dann besser auf das Problem passen (K4). Überraschend ist allerdings, dass diese Vorschläge als hilfreicher (K6) oder empfehlenswerter (K11) bewertet werden. Es ist festzuhalten, dass die Lösungen des Erfahrungs-Prompts nicht per se als schlecht bewertet wurden (siehe Tabelle 5-45). Die Lösungsvorschläge des Lösungs-Prompts waren in den entsprechenden Kategorien um ca. 0,33 höher bewertet. Es gab keine Bewertungskategorien, in denen die Lösungsvorschläge der anderen Prompts dem Lösungs-Prompt überlegen waren. Es scheint also, als würden Lösungs-Prompts, im Vergleich zu den Lösungsvorschlägen, die bei einem Erfahrungs-Prompt entstehen, besser ausgestaltete Lösungen produzieren. Dies ist nachvollziehbar, da der Empfänger beim Lösungs-Prompt eher an unterschiedliche Aspekte denkt, um eine Lösung zu beschreiben, und der Empfänger beim Erfahrungs-Prompt nur einen aus der Erfahrung heraus begründeten Vorschlag macht.

Bei der Analyse der Prompts aus Teilstudie 2.2 (Diskussionsbeiträge vorhanden) hat sich gezeigt (siehe Tabelle 5-46), dass der Kontroll-Lösungs-Prompt Lösungsvorschläge hervorbringt, die besser auf das Problem passen (K4) als der Lösungs-Prompt selbst. Der Kontroll-Lösungs-Prompt fragt nach Bezug zu den anderen Prompts. Eine Interpretation ist, dass jemand eine passendere Lösung entwickeln kann, indem derjenige andere Lösungen einbezieht, bevor er selbst eine schreibt. Andere Effekte wurden im Vergleich der zwei Prompts nicht gefunden.

Hinsichtlich der Zusammenhänge der reflexionsbezogenen Codes und der Bewertungskategorien zeigt sich, dass die Codes für Erfahrung (EXP), Wissen (KNO) und wissensbasierte Lösungsvorschläge (SUG\_KNO) einen Zusammenhang zu dem Bewertungskriterium für Glaubwürdigkeit (K7) aufweisen. Es sei darauf hingewiesen, dass alle bewerteten Lösungsvorschläge bereits erfahrungsbasiert sind (SUG\_EXP). Der Fund, dass Erfahrung eine Rolle bei der Glaubwürdigkeit spielt, bestätigt existierende Literatur zur Reflexion (siehe Abschnitte 2.4 und 2.5). Hervorzuheben ist, dass es einen Bezug von Wissen bzw. wissensbasierten Lösungsvorschlägen zu dem Kriterium der Glaubwürdigkeit gibt. Daraus lässt sich ableiten, dass der Prompt für Lösungsvorschläge generell glaubwürdige Lösungsvorschläge erzielt und es keinen Unterschied hinsichtlich der Glaubwürdigkeit gibt. Das ist zunächst überraschend, da der Austausch von persönlicher Erfahrung der Kern für kollaborative Reflexion ist und wertvoller als der Austausch von reinem Wissen betrachtet wird. Dies bedeutet, dass der Prompt bereits für den Austausch von Erfahrungen funktioniert und dahingehend nicht angepasst werden muss.



Die Analyse in Abschnitt 5.3.5.5 hat gezeigt, dass die Anzahl der Wörter eine wichtige Rolle bei fast allen Bewertungskriterien spielt. Durch die Regressionsmodelle in Abschnitt 5.3.5.2 wird dieser Einfluss quantifiziert. Zu beachten ist allerdings, dass die Anzahl von Beiträgen mit besonders wenigen Wörtern sehr niedrig ist, was die statistischen Unterschiede erklären kann. Die Ergebnisse zeigen, dass überwiegend (bis auf bei ungewöhnlichen Lösungen (K2) und Elementen, an die der Bewertende nicht gedacht hat (K3)) Lösungsvorschläge mit einer höheren Anzahl an Wörtern besser bewertete Lösungsvorschläge produzieren. Diese Ergebnisse sind verständlich, da es schwieriger ist, mit sehr wenigen Wörtern einen leicht verständlichen oder gut auf das Problem passenden Lösungsvorschlag zu formulieren. Hier kann es sich anbieten, Nutzer in Systemen zur kollaborativen Reflexion in diesem Aspekt zu unterstützen und bspw. darauf hinzuweisen, falls eine Antwort zu kurz ausfällt.

In Bezug auf die Demografie der Autoren (nicht der Bewertenden) lässt sich feststellen, dass insbesondere die demografischen Merkmale wie Land des Crowdworkers (Tabelle 5-29) und der Bildungsabschluss (Tabelle 5-28) einen Zusammenhang zu den Bewertungen der Lösungsvorschläge aufweisen. Beide Merkmale erhöhen die Bewertungen in der Hinsicht, wie gut die Lösung auf das Problem passt (K4) und inwieweit die Lösung glaubwürdig ist (K7). Das Land der Crowdworker hat höhere Bewertungen in den Kategorien der Klarheit der Formulierung (K8) und inwieweit der Lösungsvorschlag in die Praxis überführt werden kann (K9). Die nahe liegende Interpretation für diese Beobachtung ist, dass diese Personen den Lösungsvorschlag sprachlich besser vermitteln können als Personen, die Englisch (die Sprache in der Teilstudie) als Fremdsprache sprechen. Überraschenderweise werden Lösungsvorschläge derjenigen, die aus englischsprachigen Ländern stammen, als neuer bewertet. Hier gibt es keine sinnvolle Interpretation. Eventuell kann die höhere Übung mit der Sprache ein Faktor sein, einen Aspekt leichter hervorzuheben, der dann als neu bewertet wird. Hinsichtlich des Bildungsgrads werden Vorschläge von Akademikern der Kriterien, inwieweit die Lösung das Problem löst (K5) und wie hilfreich die Lösung ist (K6), höher bewertet. Die Vermutung ist hier, dass Akademiker während des Studiums eine höhere Problemlösekompetenz erreichen und daher Lösungsvorschläge passender entwickeln können. Die Erfahrung hilft lediglich bei der Beschreibung eines Lösungsvorschlags (K8). Diese Befunde konnten in den Regressionsmodellen nicht nachvollzogen werden (siehe Abschnitt 5.3.5.2). In den Regressionsmodellen zeigt sich viel mehr, dass die Erfahrung einen mittleren Effekt auf die Wahrscheinlichkeit eines als positiv bewerteten Lösungsvorschlags hat (siehe Abschnitt 0). Hier hilft die Erfahrung in Form von einer längeren Zeit, die der Autor bereits als Crowdworker gearbeitet hat, bei der Erstellung

von qualitativen Lösungsvorschlägen. Hier kann nachfolgend weiter untersucht werden, inwieweit sich diese Erfahrung qualitativ in den Lösungsvorschlägen wiederfindet.

### **5.3.6.3 Implikationen für das Design von Reflexionsunterstützung**

Tabelle 5-33 zeigt, dass die Bewertungen der Lösungsvorschläge in einem Thema ohne Diskussionsbeiträge besser bewertet waren als mit Diskussionsbeiträgen.

Es ist deutlich, dass der Lösungs-Prompt in einem Thema ohne Diskussionsbeiträge, also von Beginn eines Themas an, qualitative Lösungen hervorbringen kann. Diese Beiträge wurden auch leicht besser bewertet als Lösungsvorschläge, die innerhalb einer Diskussion entstanden sind. In weiteren Studien ist zu analysieren, ob es Prompts gibt, die innerhalb einer Diskussion besser geeignet sind, um qualitative Lösungsvorschläge zu erzeugen.

Der Kontroll-Lösungs-Prompt, der Bezug auf andere Beiträge nimmt, erzielt höhere Bewertungen im Kriterium, wie gut der Lösungsvorschlag auf das Problem passt. Dies kann eine Richtung aufzeigen: Weniger Fokus auf das Entwerfen neuer Lösungen und mehr Fokus auf das Verdichten der vorhandenen Lösungsvorschläge. Dies muss in weiteren Studien analysiert werden.

Zusätzlich zeigt Tabelle 5-44, dass der Lösungs-Prompt innerhalb der Teilstudie ohne Diskussionsbeiträge zu höher bewerteten Lösungsvorschlägen führt. Der Lösungs-Prompt ist zum Anregen von Lösungen ein probates Mittel und zeigt, dass er im Vergleich zu dem Erfahrungs- und Kontroll-Prompt zu höher qualitativen Lösungen führen kann. Es ist aber zu unterstreichen, dass die beiden anderen Prompts nicht explizit nach Lösungen fragen, sondern diese nur als Nebenprodukt hervorrufen können, bspw., weil der Proband beim Verfassen seiner Antwort noch eine Idee hatte, die er als Lösungsvorschlag transportieren wollte. Die Ergebnisse zeigen dementsprechend nicht, dass der Erfahrungs-Prompt verworfen werden sollte, sondern unterstreichen, dass der Lösungs-Prompt zu den Lösungen eine höhere Qualität hinzufügt. Diese Qualität spricht dafür, dass ein Lösungs-Prompt genutzt werden sollte, sofern es die Intention desjenigen, der die Nutzer unterstützen möchte, ist, dass Lösungen eingebracht werden.

Aufgrund der Wichtigkeit der Anzahl der Wörter kann es sinnvoll sein, den Nutzer über Prompts darauf aufmerksam zu machen, falls dieser versucht einen Lösungsvorschlag mit sehr wenigen Wörtern als Beitrag in eine Diskussion einzubringen. Der Prompt kann auch mögliche Eigenschaften eines qualitativen Lösungsvorschlags (siehe Abschnitt 5.3.2.4) beschreiben, um den Nutzer darauf aufmerksam zu machen bestimmte Aspekte in den

Text mit aufzunehmen. Das Ziel ist, dass der Empfänger dann seinen textuellen Lösungsvorschlag überdenkt und ggf. um weitere Informationen erweitert. Dies muss ebenfalls in weiteren Studien erforscht werden.

Da neue und überraschende Aspekte nicht oft bei bestimmten Prompts oder den jeweiligen Konditionen (Thema mit und ohne Diskussion) auftauchten, ist die Empfehlung dafür bestimmte Prompts zu entwerfen, die wie bei einem Brainstorming nach neuen Ideen Fragen. Diese müssen sich dann textuell von den bisherigen Lösungsprompts unterscheiden und sollten die Aspekte, dass nach etwas Neuem gefragt wird, explizit beinhalten.

### **5.3.7 Einschränkungen**

In diesem Abschnitt werden die unterschiedlichen Einschränkungen, denen die Teilstudie unterliegt, beschrieben. Hier werden nur Einschränkungen beschrieben, die nicht schon für die vorherigen Crowdsourcing-Studien in Teilstudie 2.1 und 2.2 galten.

Das Problem, dass den Crowdworkern in den Teilstudien 2.1 und 2.2 präsentiert wurde, ist relativ allgemein und nicht sehr komplex. Es ist komplizierter als eine einfache Frage, wie bspw. „Wo finde ich das FAQ?“. Das Problem ist allerdings deutlich leichter als eine Situation, die für alle Beteiligten neu wäre und nicht ad hoc mit bekannten Mitteln lösbar ist. Diese Teilstudie hat nur Aussagekraft für ein Problem, was von ähnlicher Komplexität ist, wie das vorgestellte Problem. Für Probleme anderer Komplexitätsgrade sind weitere Studien notwendig.

In der Reflexionsliteratur finden sich keine objektiven Kriterien zur Bewertung von Ideen bzw. Lösungsvorschlägen (siehe 5.3.2.4). Dementsprechend lässt sich durch die angewandten Bewertungskriterien nicht zweifelsfrei benennen, ob eine Idee für Reflexion besser oder schlechter geeignet ist. Zusätzlich gibt es die Herausforderung, dass die Lösungsvorschläge zwar von Personen aus der Domäne, den Crowdworkern, bewertet wurden, aber nicht von der Person, die das fiktive Problem hatte. Der Wert einer Idee ist erst messbar, wenn diese versucht wurde umzusetzen (Blohm u. a. 2016).

Zusätzlich hat das konstruierte Szenario die Herausforderung, dass keine echte Diskussion unter den Probanden möglich war. Somit konnten keine Lösungsvorschläge diskutiert bzw. zusammen erarbeitet werden und die Person, die den Lösungsvorschlag entwickelt hat, hatte keine Möglichkeit, Rückmeldung zu bekommen. Dies ist in der Realität ebenfalls anders. Dementsprechend haben die Qualitätsmerkmale nur Gültigkeit in Szenarien, wo Personen direkt Lösungsvorschläge einbringen, ohne diese vorher zu

diskutieren. Dies heißt, dass die Anwendbarkeit dieser Prompts ebenfalls noch in Foren bzw. Szenarien untersucht werden muss, in denen eine echte Diskussion möglich ist.

Neben der fehlenden Diskussion wurden nur Beiträge berücksichtigt, die keine expliziten Referenzen auf andere Beiträge hatten, um die Bewertung einfach zu halten. Dies ist ebenfalls eine Einschränkung, da sich in Online-Diskussionen die Themen weiterentwickeln und Beiträge aufeinander aufbauen. Allerdings gab es aus der vorherigen Teilstudie sehr wenige Beiträge, die einen erfahrungsbasierten Lösungsvorschlag (SUG\_EXP) enthalten und sich explizit auf einen anderen Beitrag beziehen.

Die Probanden haben die Bewertung lediglich anhand der Kriterien durchgeführt. Es wurden keine Beispiele oder Abgrenzungen gegeben, um die Aufgabe für die Personen kurz und durchführbar zu halten. Ebenso wurden die Bewertenden vor der Bewertung nicht in der Anwendbarkeit der Kriterien trainiert. Dementsprechend sind die Bewertungen sehr subjektiv und unterliegen nicht einem objektiven Bewertungsschema, wie bspw. das reflexionsbezogene Kodierschema (siehe Abschnitt 3.4).

Studien belegen, dass die Reihenfolge der Kategorien in einer Likert-Skala eine Auswirkung auf die Bewertung hat. Wenn die Skala von sehr hoch zu sehr niedrig verläuft, sind die Bewertungen im Schnitt höher als bei einer umgekehrten Skala (Hartley und Betts 2010). Die verwendete Skala lief von *strongly disagree* zu *strongly agree*. Dies ist zu berücksichtigen, falls die Ergebnisse der Teilstudie mit anderen Studien verglichen werden sollen.

Bei der statischen Analyse ist die Interpretation eingeschränkt, da manche Ideen vier Mal und manche Ideen fünf Mal bewertet wurden (siehe Abschnitt 5.3.3). Zusätzlich haben die Bewertenden unterschiedlich oft teilgenommen. Bei der Analyse der Bewertungskriterien und der Demografie der Autoren (nicht der Bewertenden) fließen die einzelnen demografischen Daten unterschiedlich oft ein. Eine weitere Komplexität in der statistischen Analyse ist der Fakt, dass manche Beiträge aus den vorherigen Studien, sofern sie mehrere Ideen enthielten, zwecks eindeutigerer Bewertung in mehrere Beiträge aufgeteilt wurden (siehe Abschnitt 5.3.2.3). Dadurch werden die demografischen Daten ebenfalls unterschiedlich oft einbezogen. Somit ist die Interpretation dieser Analysen eingeschränkt.

In den Regressionsmodellen waren sehr viele Variablen enthalten, was aufgrund der limitierten Datenmenge zu einer eingeschränkten Aussagekraft führt (siehe Abschnitt 5.3.5.2 und Anhang C). Ebenfalls können zwischen unabhängigen Variablen und den demografischen Variablen statistische Effekte auftauchen, da manche Ausprägungen nur

mit einer niedrigen Anzahl vorhanden sind. Dadurch sind diese Regressionsmodelle vorsichtig zu bewerten.

Die durchgeführte Teilstudie hat die Probanden auf den wissenschaftlichen Kontext aufmerksam gemacht und die Umgebung des Diskussionsforums war nur simuliert. Somit ist das Ergebnis dieser Teilstudie eingeschränkt in Systeme im privaten Umfeld oder in der Unternehmenspraxis überführbar.

### **5.3.8 Zusammenfassung**

Das Ziel der durchgeführten Teilstudie war der Vergleich von Lösungsvorschlägen anhand unterschiedlicher Bewertungskriterien, die einerseits in Themen ohne Diskussionsbeiträge (Teilstudie 2.1) und andererseits in Themen mit Diskussionsbeiträgen (siehe Teilstudie 2.2) gemacht wurden. Anhand der Literatur zu Reflexion war zu vermuten, dass eine existierende Diskussion zu einer höheren Qualität in den Lösungsvorschlägen führt.

Die Ergebnisse der Teilstudie zeigen, dass Prompts, die versuchen Lösungen anzuregen zu höher bewerteten Beiträgen führen, wenn noch keine Diskussionsbeiträge vorhanden sind. Dies scheint der gängigen Literatur von Reflexion zu widersprechen, bei der ein vorheriger Erfahrungsaustausch stattfindet, bevor ein Lösungsansatz entwickelt wird.

Zusätzlich zeigt die Teilstudie, dass der Lösungs-Prompt in Teilstudie 2.1 in unterschiedlichen Bewertungskategorien zu besseren Lösungen führt als der Erfahrung-Prompt.

Ebenfalls wird mit dieser Teilstudie deutlich, dass Akademiker und Personen, die aus englischsprachigen Ländern an der Teilstudie teilgenommen haben, in dieser Teilstudie (auch englischsprachig) in Teilen höher bewertete Lösungsvorschläge einbringen können. Dies kann daher eine interessante Zielgruppe für Prompts sein.

## **5.4 Zusammenfassende Betrachtung der Forschungsfragen**

Die Ergebnisse der einzelnen Teilstudien wurden bereits in den entsprechenden Abschnitten thematisiert. An dieser Stelle wird kurz auf die einzelnen Forschungsfragen eingegangen. Eine ausführliche Diskussion der Ergebnisse dieser Studie mit den einzelnen Ergebnissen findet sich in Kapitel 6.

Teilstudie 2.1 (siehe Abschnitt 5.1) fokussierte sich auf die Frage, inwieweit sich Erfahrungen und Lösungsvorschläge durch Prompts in einer simulierten Community of Practice Umgebung anregen lassen. Hier haben die Analysen gezeigt, dass die beiden Prompts für Erfahrungen und Lösungsvorschläge jeweils zu einem höheren Anteil von Erfahrungen und Lösungsvorschlägen führen als die jeweiligen anderen Prompts (auch im Vgl. zum Kontroll-Prompt). Sowohl der Lösungs- als auch der Erfahrungs-Prompt rufen mehr erfahrungsbasierte Lösungsvorschläge als der Kontroll-Prompt hervor. Der Kontroll-Prompt und der Lösungs-Prompt rufen mehr unbegründete Hinweise hervor als der Erfahrungs-Prompt. Somit können insbesondere Lösungs- und Erfahrungs-Prompt in vergleichbaren Szenarien eingesetzt werden.

In Teilstudie 2.2 (siehe Abschnitt 5.2) wurden zwei Fragen behandelt. Einerseits wurde untersucht, inwieweit feinere Nuancen in den Formulierungen von Prompts eine Rolle beim Anregen von Lösungsvorschlägen für kollaborative Reflexion spielen. Hier zeigten sich keine messbaren Unterschiede in der Häufigkeit der Lösungsvorschläge in den Texten der Probanden, was impliziert, dass sich feinere Unterschiede in den Prompts in der Teilstudie nicht auf die Beiträge ausgewirkt haben. Andererseits wurde die Frage thematisiert, inwieweit das Vorhandensein von Diskussionsbeiträgen einen Einfluss auf die Erstellung von Lösungsvorschlägen in einer simulierten Community of Practice Plattform hat. Die Ergebnisse in der Studie zeigen, dass existierende Diskussionsbeiträge nicht häufiger zu Lösungsvorschlägen führen (im Vgl. zu einem Thema ohne Diskussionsbeiträge), wenn den Nutzern ein Lösungs-Prompt gezeigt wird.

Die dritte Teilstudie (2.3 in Abschnitt 5.3) hat untersucht, inwieweit sich die Qualität von Lösungsvorschlägen unterscheidet, wenn Lösungs-Prompts in Diskussionen mit und ohne Diskussionsbeiträgen genutzt werden. Die Ergebnisse zeigen, dass die Lösungsvorschläge aus Teilstudie 2.1 leicht besser bewertet wurden, wenn keine Diskussionsbeiträge vorhanden waren. Regressionsmodelle zeigen, dass die fehlenden Diskussionsbeiträge nicht der einzige Faktor waren, und andere Faktoren, wie bspw. die Erfahrung des Autors, ebenfalls einen Einfluss auf die Bewertung hatten.

## **6 Diskussion der Ergebnisse beider Studien**

In diesem Kapitel werden die Ergebnisse der unterschiedlichen Studien dieser Arbeit zusammengefasst, integriert betrachtet und hinsichtlich des Konzepts diskutiert. Weiterhin wird die Beantwortung der Forschungsfrage thematisiert. Zusätzlich werden die Einschränkungen der Arbeit diskutiert.

### **6.1 Diskussion der Ergebnisse**

Die Zielsetzung der Arbeit war zu untersuchen, inwieweit sich Aspekte kollaborativer Reflexion in Communities in einem beruflichen Kontext durch Prompts anregen lassen. Zur Untersuchung des Konzepts (siehe Kapitel 3) wurden Prompts für kollaborative Reflexion in zwei Umgebungen untersucht: Zum einen wurde in Studie 1 (siehe Kapitel 4) der Einsatz von Prompts in einer Feldstudie in einer Community of Practice in öffentlichen Verwaltung getestet. Zum anderen wurde eine Serie von drei Studien durchgeführt (siehe Abschnitte 5.1, 5.2 und 5.3), in der fokussiert der Effekt von Prompts in einer Crowdsourcing-Umgebung evaluiert wurde. In diesem Abschnitt werden die Ergebnisse dieser beiden Untersuchungslinien zusammengefasst diskutiert. Auf diese Weise wird versucht, die Forschungsfrage differenziert zu beantworten.

#### **6.1.1 Häufigkeit von Beiträgen**

In Studie 1 (siehe Kapitel 4) wurden Prompts zur Unterstützung kollaborativer Reflexion in zwei unterschiedlichen öffentlichen Verwaltungen, die als Arbeitsagentur beide den gleichen Kontext haben, eingesetzt. Die Mitarbeiter in beiden Organisationen hatten das Bedürfnis nach mehr Austausch über die eigene Praxis untereinander. Dazu wurde eine Community of Practice Plattform entwickelt und mit Prompts als Funktionalität zur Unterstützung von Reflexion versehen.

Die Ergebnisse zeigen, dass systemgenerierte Prompts nicht zu einem Unterschied in der Häufigkeit der Antworten geführt haben. Dies impliziert, dass sich Prompts in der durchgeführten Studie nicht dazu geeignet haben, die Empfänger dazu zu motivieren mehr Antworten zu verfassen. Dies liegt bei den Prompts zu Erstellung neuer Themen daran, dass die Prompts unterhalb der Liste der Themen angezeigt wurden (in dem Bereich in dem Nutzer die Beiträge verfassen) und die Nutzer dort erst hin navigieren mussten. Dementsprechend haben die Prompts in dieser Studie eher beim Verfassen von Beiträgen als beim Anregen neuer Beiträge unterstützt. Rückblickend behinderte die Positionierung möglicherweise den Einfluss der Prompts auf die Antwortrate. Zusätzlich ist die Kultur in beiden Unternehmen ein Faktor, da die zuständigen Teammanager die

Nutzung der Community of Practice Plattform nicht offiziell unterstützt haben. Die Nutzung wurde toleriert aber nicht aktiv gefördert. Dies führte möglicherweise dazu, dass die Mitglieder die Plattform eher mit der Absicht genutzt haben, Beiträge lesen zu wollen, anstatt zeitaufwendig selbst Inhalte beizutragen. Diese niedrige Antwortrate bei Prompts findet sich auch in der Literatur (Renner u. a. 2016). Bei den drei Studien, die in der Crowdsourcing-Umgebung durchgeführt wurden, spielt die Antwortrate der Probanden keine Rolle, da in der Crowdsourcing-Plattform nicht überprüft werden konnte, wie viele Personen den Crowdsourcing-Job angesehen aber nicht durchgeführt haben. Zusätzlich wurden die Teilnehmer für Beiträge direkt kompensiert und die Situation unterscheidet sich dahingehend von der Situation in Studie 1, in der die Teilnehmer die Plattform während ihrer Arbeit freiwillig nutzen konnten.

### **6.1.2 Nutzergenerierte Prompts**

Bei nutzergenerierten Prompts ist in dem Anwendungsfall in Slowenien eine höhere Antwortrate aufgetreten (siehe Abschnitt 4.4.6). Nutzergenerierte Prompts stellen für Nutzer eine Möglichkeit dar, bei Erstellung eines neuen Themas eine eigene Frage an dieses Thema anzuhängen, welche dann visuell für die anderen Nutzer hervorgehoben wird und als Prompt angezeigt wird. Ein Grund für die teilweise höhere Antwortrate kann eine vertrautere Formulierung für den gegebenen Kontext in dem Thema sein, welcher von den übrigen systemgenerierten Prompts nicht erreicht wird. Ein anderer Grund kann sein, dass der Text des Nutzers doppelt dargestellt wird: sowohl regulär formatiert innerhalb des Beitrags und visuell hervorgehoben als Prompt. Somit lesen Empfänger den Text häufiger. Der Text wurde doppelt dargestellt, da nicht alle Nutzer Prompts gesehen haben. Dennoch kann diese doppelte Darstellung ebenfalls der Grund für die höhere Antwortrate sein. Aufgrund der Durchführung der Studie lässt sich ein Unterschied in der Analyse nicht herausarbeiten. Hier sind weitere Studien notwendig, um zu ermitteln, welcher Faktor der ausschlaggebende Faktor ist. Personen, die aktuell keine Prompts angezeigt bekommen haben, haben auf diese Weise immerhin die Frage innerhalb des regulären Textes lesen können. Deren Antworten sind aber nicht während eines Prompts erfolgt und wurden daher in der inhaltlichen Auswertung nicht berücksichtigt. Diese Art der Darstellung wurde absichtlich so gewählt, damit weiterhin eine Diskussion möglich ist. Nutzer wären verwundert, falls sie mit Kollegen sprechen und im Gespräch auffällt, dass nicht bei jedem Kollegen die Frage sichtbar ist.

Nutzergenerierte Prompts können ein probates Mittel darstellen, um zu versuchen die Antworten innerhalb von Themen, also bei der Erstellung von Antworten bzw. Diskussionsbeiträgen zu erhöhen. Zusätzlich haben die Analysen ergeben, dass die Texte,



die durch nutzergenerierte Prompts beeinflusst wurden, häufiger in Zusammenhang mit erfahrungsbasierten Lösungsvorschlägen (SUG\_EXP) stehen. Da dies ein wichtiges Element in der Reflexion ist (Woerkom und Croon 2008; Raelin 2002; Dyke 2006), stellen nutzergenerierte Prompts eine interessante Richtung für künftige Studien dar. In Abschnitt 7.2 wird das Layout der nutzergenerierten Prompts diskutiert und ein Designvorschlag für künftige Studien erarbeitet.

### **6.1.3 Effekte von Prompts**

In Studie 1 (siehe Kapitel 4) wurden die Texte zu 40-46 % als durch Prompts beeinflusst gewertet. Das ist ein relativ hoher Wert und impliziert, dass die Prompts einen Effekt auf die Texte der Nutzer haben können. Es ist zu beachten, dass die Beurteilung nicht von den Autoren der Beiträge selbst stammt. Dementsprechend kann eine Übereinstimmung zwischen Prompt und Text zufällig sein, bspw., weil der Autor ohnehin den bestimmten Inhalt in seinem Beitrag verfassen wollte. Dennoch impliziert dieser Wert, dass Prompts in der Praxis in einer Community of Practice Plattform in der täglichen Arbeit genutzt werden können. Analysen zeigen bspw., dass wenn ein Prompt in dieser Studie beachtet wurde, der Beitrag häufig mit den Codes für Erfahrung (EXP), Hinweise (ADV), oder erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) korreliert (siehe Tabelle 4-19). Die meisten Studien konzentrieren sich bei Prompts eher auf Studien im Bildungsbereich, bei denen Prompts den Empfängern bei der Erledigung einer (oftmals pflichtmäßigen bzw. benoteten Aufgabe) angezeigt wird. Studien, die Prompts hinsichtlich kollaborativer Reflexion in beruflichen Situationen analysieren, sind selten. Dementsprechend kann dieser Fund nicht in einen größeren Kontext gesetzt werden und sollte in weiteren Studien ggf. in anderen Domänen (bspw. in Unternehmen, die keine öffentliche Verwaltung sind) reproduziert werden.

Aus der ersten Studie in der Crowdsourcing-Umgebung zeigt sich, dass die Prompts für Erfahrungen und Lösungen einen Zusammenhang zu den entsprechenden reflexionsbezogenen Codes für Erfahrungen (EXP) und Lösungsvorschläge (SUG) haben. Das zeigt, dass sich die Ziele für Erfahrung (Z6) oder Lösungsvorschläge (Z8) durch Prompts unterstützen lassen. Zusätzlich haben Lösungs- und Erfahrungs-Prompts im Vergleich zum Kontroll-Prompt häufiger erfahrungsbasierte Lösungsvorschläge (SUG\_EXP) hervorgerufen. Die Ergebnisse zeigen, dass in dieser Umgebung Prompts für Reflexion einen Effekt auf die Texte der Empfänger haben. Die beobachteten Effekte sind deutlich stärker als in der Feldstudie. In der Feldstudie haben insbesondere die nutzergenerierten Prompts einen Zusammenhang zu erfahrungsbasierten Lösungsvorschlägen gezeigt. In anderen Studien haben Prompts für Lösungen im

Vergleich zu anderen Prompts nicht häufiger zu Lösungsvorschlägen geführt (Renner u. a. 2016), was allerdings den Grund haben kann, dass Renner et al. bei den übrigen Prompts ebenfalls nach Lösungen gefragt haben. Bei Renner et al. wird ein *Problem-Prompt* mit dem Text „*What changes would you suggest for similar situations?*“ (Renner u. a. 2016, 3) verwendet und die daraus resultierenden Texte können als Lösungsvorschläge gewertet sein. Somit passt das bei Renner et al. gefundene Ergebnis sehr gut zu dem in Teilstudie 2.2 gefundenen Ergebnis (siehe unten), wonach sich Prompts mit leicht unterschiedlichen Formulierungen nicht in der Wirkung unterscheiden.

Es wurde bereits gezeigt, dass Crowdsourcing-Umgebungen mit Laborstudien anderer Studien vergleichbar sind (siehe Abschnitt 3.6.6). Die Probanden hatten eine klare Aufgabe in der Studie und dementsprechend sind diese Ergebnisse nicht direkt mit Praxissituationen, wie mit der Community of Practice-Umgebung in Studie 1, vergleichbar. Dennoch können diese Ergebnisse als Vorbereitung für die Evaluierung in der Praxis betrachtet werden und sie bieten einen begründeten Ausgangspunkt für weitere Arbeiten in diese Richtung.

Die zweite Crowdsourcing-Studie (Teilstudie 2.2; siehe Abschnitt 5.2) zeigt, dass Prompts deren Formulierungen nahe beieinander liegen, sich in der Wirkung nicht signifikant unterscheiden. Die in der Teilstudie 2.1 beobachteten Effekte, in denen die unterschiedlichen Prompts unterschiedliche Wirkung auf die reflexionsbezogenen Codes haben, wurden nicht beobachtet (siehe Abschnitt 5.1). Interessanterweise war in dieser Studie der Kontroll-Lösungs-Prompt nicht schlechter, was die Häufigkeit der reflexionsbezogenen Codes betrifft. Ein anderer Grund kann sein, dass es für die Teilnehmer leichter war, einen Lösungsvorschlag zu formulieren, da die angezeigten Kommentare in Teilstudie 2.2 selbst Lösungsvorschläge enthalten haben. Dies kann einen Einfluss auf die Lösungsvorschläge der Teilnehmer gehabt haben. Dieser Prompt hatte den Bezug zu anderen Prompts zum Gegenstand. Dementsprechend lässt sich ableiten, dass bei deutlich unterschiedlichen Prompts eine Wirkung beobachtet werden kann und dass scheinbar die Empfänger die Prompts nicht so genau lesen, sodass Feinheiten in der Formulierung keinen Effekt haben. Eine andere Interpretation ist, dass die Probanden die Konzepte der Prompts (expliziter Bezug auf die anderen Kommentare) in dem Moment als unpassend betrachtet haben. Ein Ansatz, um dem Problem unpassender oder nicht ausreichend deutlicher Formulierungen zu entgehen, sind Workshops mit potenziellen Nutzern, um gemeinsam passende Prompts für Reflexion zu erarbeiten (Kocielnik u. a. 2018).

In Studie 1 korrelierte der Prompt „*What have you learned from this discussion so far?*“ (siehe Tabelle 4-6 und Tabelle 4-16; Z9, Z10) in dem Anwendungsfall in Slowenien mit Lernen (D\_LOOP) und im Anwendungsfall in Kroatien mit den Codes für Zustimmung (AGR) und Ablehnung (DISAGR). Alle drei Codes sind positiv für eine Diskussion. Dennoch ist die Wirkung ambivalent und weitere Studien sind notwendig, um die Wirkung des Prompts zu prüfen. Aus der Studie heraus lassen sich allerdings wenige Aussagen hinsichtlich der Formulierung von Prompts ableiten, da für die meisten Prompts, aufgrund der niedrigen Aktivität der Nutzer, nicht ausreichend Daten gesammelt werden konnten, um eine aussagekräftige Analyse durchzuführen. Allerdings wurde im Anwendungsfall in Kroatien ein Text nur in einem Fall als übereinstimmend mit diesem Prompt bezeichnet. Die Wirkung des Prompts wurde daher nur in einem Anwendungsfall beobachtet. Dementsprechend lassen sich diese Ziele (Z9, Z10) eingeschränkt (nur in einem von zwei Anwendungsfällen) durch Prompts unterstützen. Mehr Daten sind notwendig, um die Wirkung dieses spezifischen Prompts zu untersuchen.

Inwieweit alle übrigen Ziele durch Prompts unterstützt werden können, kann aufgrund der geringen Datenmenge in Studie 1 (siehe Kapitel 4) nicht eindeutig bewertet werden. Hier sind weitere Studien notwendig.

In der Teilstudie 2.3 wurden die erfahrungsbasierten Lösungsvorschläge (SUG\_EXP) der Teilstudien 2.1 und 2.2 bewertet. Wenn man die Qualität der Lösungsvorschläge aus Teilstudie 2.1 resultierend aus Lösungs-Prompt bzw. Erfahrungs-Prompt vergleicht, dann ist festzustellen, dass die Texte der Lösungs-Prompts in mehreren Kategorien besser bewertet sind: der Lösungsvorschlag passt zum Problem (K4), er löst voraussichtlich das Problem (K5), er ist hilfreich (K6), er ist in die Praxis überführbar (K9), er hat vermutlich einen Einfluss auf das Verhalten des Bewertenden (K10) und der Bewertende würde ihn weiterempfehlen (K11). Bei den anderen Kriterien gab es keinen Unterschied. Tabelle 5-45 zeigt, dass der Durchschnittswert der Bewertungen der Texte von Erfahrungs-Prompt ca. 0,33 unter dem der Lösungs-Prompts liegt. Dieser Wert ist relativ niedrig, sodass Erfahrungs-Prompts genutzt werden können, wobei die teilweise entstehenden Lösungs-Prompts an sich nicht qualitativ als minderwertig anzusehen sind. Dementsprechend lassen sich in einem neuen Thema ohne Diskussionsbeiträge beide Prompts sehr gut nutzen. Es sind weitere Studien notwendig, um zu evaluieren, inwieweit auch andere Prompts genutzt werden könnten, um ähnliche Ergebnisse hervorzurufen.

Ein Effekt von Over-Prompting (siehe Abschnitt 2.10) konnte in Studie 1 nicht beobachtet werden (Nückles, Hübner und Renkl 2006). In dem einen Anwendungsfall in Studie 1 wurden Prompts nach Ablauf des Studienzeitraums konsequent angezeigt, wohingegen in dem anderen Anwendungsfall nach Ablauf des Studienzeitraums keine

Prompts angezeigt wurden. Die Antwortraten verändern sich zwar, wenn man in beiden Anwendungsfällen die ursprünglichen Studienzeiträume im Vergleich zu dem gesamten Datenerfassungszeitraum vergleicht, wobei die Aktivität der Nutzer über Zeit abgenommen hat. Diese Abnahme der Aktivität macht es unattraktiv neue Themen beizutragen, weil weniger Leute diese Beiträge lesen. Dementsprechend scheint Over-Prompting in den beiden Studien nicht vorhanden zu sein. Somit scheint der kurze Zeitraum von drei Monaten, in denen die Nutzer Prompts gesehen haben, in Verbindung mit der niedrigen Nutzungsrate der Plattform noch nicht zu Over-Prompting zu führen. Da die Nutzer aber nicht zu den Prompt-Texten befragt wurden, ist offen, inwieweit sie sich womöglich zu oft wiederholt haben, was ein Problem beim Design von Prompts ist (Kocielnik u. a. 2018). Hier ist wichtig, dass in wissenschaftlichen Studien die Menge der Prompts sorgfältig gewählt ist. Auf der einen Seite müssen die Prompts ausreichend abwechslungsreich sein, damit Ermüdungserscheinungen bei den Nutzern vermieden werden und auf der anderen Seite muss eine ausreichende Menge an Daten erhoben werden, damit der Prompt evaluiert werden kann.

#### **6.1.4 Zeitpunkt von Lösungs-Prompts in der Reflexion**

Die ersten beiden Teile der Crowdsourcing Studie wurden so angelegt, dass der Lösungs-Prompt in beiden Studien identisch ist. Der verbleibende Unterschied zwischen den beiden Studien ist nur die Abwesenheit (Teilstudie 2.1) bzw. das Vorhandensein (Teilstudie 2.2) von Diskussionsbeiträgen in dem Thema. Auf diese Weise kann geprüft werden, ob der Zeitpunkt, wann der Prompt gezeigt wird, einen Einfluss auf die Anzahl der Lösungsvorschläge hat, die der Lösungs-Prompt erzeugt. In Teilstudie 2.1 wurden diese direkt von Anfang an in einem leeren Thema gezeigt, in Teilstudie 2.2 hingegen, in einem Thema, zu dem Personen mit Antworten beigetragen haben. Die Analyse hat gezeigt (siehe Abschnitt 5.2.5), dass kein Unterschied existiert und der Lösungs-Prompt in beiden Bedingungen ähnlich viele Lösungsvorschläge erzeugt.

Kollaborative Reflexion wird in der Literatur als Gruppenprozess beschrieben, der teils individuelle Aktivitäten hat (Prilla, Pammer und Krogstie 2013). Dieser Prozess ist durch die Elemente wie Erfahrungsaustausch, Diskussion, und dem Hinterfragen von Annahmen, sowie dem Generieren, Diskutieren und Verwerfen von Ideen geprägt (Daudelin 1996; Høyrup 2004; Woerkom und Croon 2008; Raelin 2002; Dyke 2006; Hatton und Smith 1995). Es wird in der Literatur impliziert, dass die Teilnehmer kollaborativer Reflexion zunächst diskutieren, eigene Erfahrungen austauschen, am Ende Lösungen diskutieren und deren Umsetzung planen. Dementsprechend ist der Fund aus dem Vergleich des Lösungs-Prompts aus den Teilstudien 2.1 und 2.2 überraschend. Es

gab keinen Unterschied in der Frequenz der Lösungsvorschläge. Um zu klären, ob es Unterschiede in der Qualität der Lösungsvorschläge gibt, wurde Teilstudie 2.3 durchgeführt (siehe Abschnitt 5.3).

Teilstudie 2.3 hatte zum Ziel die Qualität der Lösungsvorschläge bewerten zu lassen. Alle erfahrungsbasierten Lösungsvorschläge aus den Teilstudien 2.1 und 2.2 wurden durch Crowdworker (als Experten des Problems, welches in den Teilstudien 2.1 und 2.2 diskutiert wurde) bewertet. Bezogen auf den Lösungs-Prompt zeigt Tabelle 5-33, dass die Lösungsvorschläge aus einem neuen Thema ohne Antworten (Teilstudie 2.1) im Vergleich zu einem Thema mit Diskussionsbeiträgen (Teilstudie 2.2) hinsichtlich der folgenden Kriterien besser bewertet wird: Passung des Lösungsvorschlags auf das Problem (K4), inwieweit der Lösungsvorschlag das Problem löst (K5), inwieweit er hilfreich ist (K6), Glaubwürdigkeit (K7), Klarheit der Beschreibung (K8), Möglichkeit ihn in die Praxis zu überführen (K9), inwieweit er einen Einfluss auf die Handlungen des Bewertenden hat (K10), inwieweit er weiterempfohlen wird (K11) und ob der Bewertende ihn nutzen würde, falls er das Problem selbst hat (K12). Dementsprechend ist der Lösungs-Prompt in einer Situation eines neuen Themas ohne Antworten besser geeignet als in einer Situation mit Diskussionsbeiträgen. Eine Vermutung für den aktuellen Unterschied ist, dass die vorhandenen Lösungsvorschläge die Teilnehmer dazu verleitet haben, die eigenen Lösungsvorschläge knapper zu halten, bspw., weil ihre eigenen Lösungsvorschläge die existierenden Lösungsvorschläge wiederholt hätten. Da der Unterschied allerdings sehr gering ist (siehe Tabelle 5-34), lässt sich der Lösungs-Prompt dennoch gut in beiden Situationen nutzen. Dieses Ergebnis ist überraschend, da zu erwarten ist, dass die Diskussion förderlich für die Qualität neuer Vorschläge ist.

Die Ergebnisse von Teilstudie 2.3 zeigen, dass eine Diskussion, die aus mehreren anderen Lösungsvorschlägen besteht (siehe Abbildung 5-6), der Qualität neuer Lösungsvorschläge nichts Neues hinzufügt und hinsichtlich der zuvor erwähnten Kriterien negativer ausgelegt werden kann. Dieses Ergebnis deutet darauf hin, dass existierende Modelle von Reflexion dahingehend angepasst werden sollten, dass das Entwickeln von neuen Lösungsvorschlägen gemeinsam ohne Erfahrungsaustausch von Anfang an möglich ist. Dies bedeutet, dass in kollaborativer Reflexion direkt nach Teilen der Problemstellung oder der Herausforderung mit der Entwicklung von Lösungen begonnen werden kann, ohne dass zunächst Erfahrungen ausgetauscht werden. Dementsprechend können Prompts für Lösungsvorschläge explizit von Anfang an genutzt werden, ohne dass dies einen negativen Einfluss auf die Häufigkeit oder die Qualität von Lösungsvorschlägen hat. Eine andere Interpretation kann sein, dass die abgebildete Diskussion nicht ausreichend war, um höherqualitative Lösungsvorschläge zu entwickeln.

Für die Interpretation der Ergebnisse muss allerdings die Natur der Kommentare, welche die Diskussion in Teilstudie 2.2 abbilden (siehe Abbildung 5-6), beachtet werden. Alle drei Kommentare sind sehr konstruktiv und sehr lösungsorientiert. Es gibt keine Kommentare, die ausschließlich unbegründete Hinweise (ADV) liefern oder nur Erfahrungen mitteilen (EXP), ohne einen Lösungsvorschlag zu machen (die Beiträge in dieser Teilstudie liefern aber erfahrungsbasierte Lösungsvorschläge). Dementsprechend haben die beschriebenen Ergebnisse nur Gültigkeit für diese Situationen.

### **6.1.5 Anregung von neuartigen oder ungewöhnlichen Ideen**

Bei der Bewertung der Lösungsvorschläge in Teilstudie 2.3 wurden die Ideen nach den folgenden Kriterien evaluiert, um zu testen, inwieweit bei dem Prompt neuartige Ideen produziert werden: Neuheit (K1), ungewöhnliche Aspekte des Lösungsvorschlags (K2) und Elemente, an die der Bewertende nicht gedacht hätte (K3).

Bei einem Vergleich hinsichtlich der Lösungs-Prompts gab es für diese drei Kategorien in den Konditionen für neue Themen ohne Diskussionsbeiträge (Teilstudie 2.1) und Themen mit Diskussionsbeiträgen (Teilstudie 2.2) keine Unterschiede. Dies bedeutet, dass die Lösungsvorschläge in beiden Situationen als ähnlich neu bzw. ungewöhnlich bewertet werden.

Die Analyse hat gezeigt, dass die durchschnittlichen Werte für diese drei Kategorien am niedrigsten waren und alle unter dem Wert drei liegen, was bedeutet, dass die Bewertenden am häufigsten die Lösungsvorschläge als nicht neu oder ungewöhnlich bewertet haben (siehe Tabelle 5-31). Hier muss beachtet werden, dass die Bewertenden alle die höchste Erfahrungsstufe auf der Crowdsourcing-Plattform haben, und dass daher die Neuheit der Ideen für diesen Personenkreis nicht gegeben ist. Das Ergebnis impliziert, dass sich der Lösungs-Prompt in der aktuellen Formulierung bei der Probandengruppe nicht eignet (sowohl bei den Autoren und bei den Bewertenden waren nur Probanden der höchsten Erfahrungsstufe der Crowdsourcing-Plattform enthalten), um neue Lösungen anzuregen, sondern eher um Lösungsvorschläge basierend auf bekannten Lösungen weiterzugeben. Für den praktischen Einsatz solcher Prompts sind weitere Studien notwendig. Dazu muss auch unterschieden werden, zu welchem Zweck die Prompts eingesetzt werden sollen. Wenn bspw. die Zielsetzung das Entwickeln gänzlich neuer Lösungsvorschläge ist, dann eignet sich der Prompt im vorliegenden Szenario nicht. Eine Möglichkeit für weitere Forschung in diese Richtung könnte das Umformulieren des Prompts hinsichtlich neuer Ideen oder eine Kombination von Kreativitätstechniken und Prompts sein, um gänzlich neue Ideen zu entwickeln.

### 6.1.6 Unerwartete Erlebnisse als Gegenstand in Prompts

Die Prompts in Studie 1, die bspw. nach unerwarteten Erlebnissen in den letzten zwei Wochen gefragt haben, haben nie eine Antwort bekommen. In der Literatur werden unerwartete Erlebnisse als ein zentraler Auslöser für Reflexion beschrieben, welcher den Beginn einer individuellen oder kollaborativen Reflexion darstellt (Dewey 1933; Schön 1983; Boud, Keogh und Walker 1985). Dementsprechend ist dieses Ergebnis überraschend. Eine Interpretation ist, dass sich die Empfänger, in den Situationen, in denen sie die Prompts empfangen, nicht an solche Ereignisse erinnern bzw. diese in dem Moment nicht als wichtig genug erscheinen. Die fehlende Beachtung dieses Prompts kann der Umgebung der Community of Practice Plattform, der Kultur in den Organisationen oder auch der Kultur der Region geschuldet sein. Es ist festzuhalten, dass diese Prompts nur in ca. 55-57 % der Fälle gesehen wurden, was der niedrigste Wert für alle Prompts ist (siehe Tabelle 4-6 und Tabelle 4-16). Im Verhältnis zu anderen Prompts waren diese Prompts weniger sichtbar, aber hinsichtlich der absoluten Häufigkeit waren die Prompts dennoch in ca. 150 – 200 Seitenaufrufen für Nutzer sichtbar.

Dies bedeutet, dass zwar Unerwartetes an sich ein Auslöser für Reflexion ist, Prompts aber einen anderen Weg finden müssen, um Reflexion anzuregen. Eine Möglichkeit kann sein, dass Prompts versuchen erst eine private Notiz anzuregen, über die man später selbst reflektieren möchte, um dann in einem späteren Schritt bzw. Prompt das Teilen der Notiz anzuregen. Auf diese Weise hätte der Nutzer Zeit, sich mit der Herausforderung gedanklich auseinanderzusetzen und teilt diese Gedanken nicht ad hoc. Eine weitere Möglichkeit kann sein, dass die Prompts nicht nach Unerwartetem in den eigenen Erfahrungen fragen, sondern den Fokus leicht verschieben auf aktuelle Herausforderungen in der Praxis der Arbeitsgruppe. Möglicherweise fällt es so leichter, über eine Problemsituation zu sprechen als direkt auf eigene Herausforderungen hinzuweisen, als direkt auf eigene Herausforderungen hinzuweisen, die von anderen als mögliche Unzulänglichkeiten eigener Fähigkeiten interpretiert werden können.

Die Prompts hinsichtlich unerwarteten Erlebnissen wurden ausschließlich auf der Startseite der Community of Practice Plattform genutzt (siehe Diskussion in Abschnitt 6.1.9.2). Dementsprechend ist hier nicht eindeutig abzuleiten, ob diese Prompts aufgrund ihres Inhalts nicht beantwortet wurden, oder ob die fehlenden Antworten daran lagen, dass Nutzer keine Beiträge über die Startseite beitragen wollten. Hier sind weitere Studien notwendig, um die Anwendbarkeit dieser Prompts zu evaluieren.

### **6.1.7 Anzahl der Wörter als Faktor**

In den Studien in der Crowdsourcing-Umgebung wurde festgestellt, dass die Anzahl der Wörter immer einen Zusammenhang zu den reflexionsbezogenen Codes (siehe Abschnitt 5.1.5 und 5.2.5) bzw. den Bewertungen der Lösungsvorschläge (siehe Abschnitt 5.3.5.5) hat. Die Ergebnisse in allen drei Teilstudien in Studie 2 zeigen, dass es bei sehr wenigen Wörtern wesentlich seltener zu einem reflexionsbezogenen Code oder einer guten Bewertung der Lösung kommt. Umgekehrt wurden wenig Korrelationen von einer hohen Anzahl an Wörtern (81-120) zu der höchsten Anzahl an Wörtern (121+) gefunden. Die Ergebnisse lassen sich dahingehend interpretieren, dass es sehr wichtig ist, eine minimale Anzahl an Wörtern in einen Beitrag einfließen zu lassen, der in eine Erfahrung oder einen Lösungsvorschlag münden soll. Ab einer Anzahl von 81 Wörtern scheinen weniger neue Codes hinzugefügt zu werden bzw. die Bewertung der Lösungsvorschläge bleibt gleich. Zusätzlich hat der Lösungs-Prompt durchschnittlich mehr Wörter als der Erfahrungsprompt in Studie 1 oder der Kontroll-Lösungs-Prompt in Teilstudie 2.1 hervorgerufen (beides siehe Abschnitt 5.3.5.5), was in der Literatur bestätigt wird (Renner u. a. 2016). Hier muss auch beachtet werden, dass Beiträge, die zu lang sind (> 500 Wörter) nicht mehr gelesen werden (Peters und Hewitt 2010). Eine andere Quelle nennt eine Länge von Beiträgen mit 100 bis 200 Wörtern als moderate Länge (Wise, Zhao und Hausknecht 2013). Vor dem Hintergrund sind die in dieser Studie verfassten Beiträge hinsichtlich der Länge weiterhin im Rahmen. Es kann in weiteren Studien versucht werden, auf zu kurze Beiträge aufmerksam zu machen. Es sollte dabei berücksichtigt werden, dass dies dazu führen könnte, dass Beiträge zu lang werden. Dementsprechend sollten in einer Studie Mechanismen für beide Fälle (zu kurze und zu lange Beiträge) evaluiert werden.

Ein Vergleich mit der Studie 1, die in einer realen Praxisumgebung durchgeführt wurde, ist nicht direkt möglich, da unterschiedliche Sprachen genutzt wurden, die zum Transport ähnlicher Ideen ggf. mehr oder weniger Wörter nutzen.

Eine Idee ist, die Ergebnisse hinsichtlich der Anzahl der Wörter in künftigen Prompts umzusetzen, um Nutzern aufzuzeigen, dass bspw. ein Beitrag zu knapp ist. Ein erster Entwurf dazu findet sich in Abschnitt 7.2.

### **6.1.8 Mögliche Einflüsse der demografischen Daten**

In den Crowdsourcing-Studien wurden unterschiedliche Zusammenhänge zwischen den reflexionsbezogenen Codes bzw. der Bewertung der Lösungsvorschläge und unterschiedlichen demografischen Daten gefunden. Hauptsächlich interessant scheinen die Merkmale des Landes des Crowdworkers und des Bildungsabschlusses des



Crowdworkers zu sein. Bei dem Land der Crowdworker kann angenommen werden, dass die Personen durchschnittlich eine höhere Sprachfertigkeit der lokalen Sprache haben als Personen anderer Länder. Es wird ebenfalls angenommen, dass dies bei den Probanden in dieser Studie der Fall ist. Die Sprache der Crowdsourcing-Studien war Englisch.

Die Ergebnisse in Teilstudie 2.1 zeigen (siehe Abschnitt 5.1.4), dass Personen aus englischsprachigen Ländern einen signifikant höheren Anteil der Hinweise (ADV) und erfahrungsbasierten Lösungsvorschläge (SUG\_EXP) verfasst haben. In Teilstudie 2.2 (siehe Abschnitt 5.2.4) haben englischsprachige Personen signifikant mehr Erfahrungen (EXP) und erfahrungsbasierte Lösungsvorschläge eingebracht. Die beiden Studien zeigen somit, dass Personen, die eine höhere Sprachfertigkeit haben, leichter Erfahrungen bzw. erfahrungsbasierte Lösungsvorschläge und Hinweise einbringen können. Durch die Bewertung der Lösungsvorschläge aus Teilstudie 2.3 (siehe Abschnitt 5.3.6) ist deutlich, dass diese in den folgenden Kriterien besser bewertet werden: Neuheit (K1), Grad, wie sehr der Lösungsvorschlag zum Problem passt (K4), Glaubwürdigkeit (K7), Klarheit der Formulierung (K8) und wie sehr der Bewertende meint den Lösungsvorschlag des Autors in die Praxis überführen zu können (K9).

Hinsichtlich des Bildungsgrads gibt es keine Unterschiede in der Häufigkeit, wie oft ein reflexionsbezogener Code auftritt. Es gibt jedoch mehrere Unterschiede hinsichtlich der Qualität der Lösungsvorschläge. Lösungsvorschläge dieser Personengruppe werden in den folgenden Kategorien höher bewertet: Wie sehr das Problem auf die Lösung passt (K4), inwieweit die Lösung das Problem lösen kann (K5), wie hilfreich der Lösungsvorschlag ist (K6) und ob der Lösungsvorschlag glaubwürdig ist (K7). Es zeigt sich, dass der Bildungsgrad einen höheren Einfluss auf die Qualität der Lösungsvorschläge hat als bspw. die Erfahrung des Autors.

In der Feldstudie (Studie 1) wurden keine demografischen Daten der Benutzer erhoben.

Die Daten zeigen, dass bestimmte demografische Daten keinen Einfluss haben (bspw. Alter, Geschlecht), andere Daten, wie die Sprache und der Bildungsgrad, jedoch schon. Hier ist eine Möglichkeit, Prompts eher diesen Personen zu zeigen, um die Qualität der Diskussion zu fördern, falls die Daten bekannt sind. Damit unterscheiden sich diese Ergebnisse von anderen Studien, in denen bspw. Prompts bei männlichen Teilnehmern besser funktioniert haben (van Oldenbeek u. a. 2019).

### **6.1.9 Diskussion des Layouts auf der Community-Plattform**

Das Design der Prompts in Studie 1, die im praktischen Alltag in einer Community of Practice Plattform durchgeführt wurde, war nicht ideal. Dieses wird in diesem Abschnitt kritisch diskutiert.

#### **6.1.9.1 Design der Prompts**

Das Design der Prompts hat sich bei Studie 1 bei den jeweiligen Organisationen in das allgemeine Design und das Farbschema der Plattform eingefügt. Die Farbschemata waren jeweils am Logo der jeweiligen Organisation ausgerichtet. Auf diese Weise waren die Prompts nicht so sichtbar, wie sie hätten sein können. Hier ist es wichtig, einen Kompromiss zu finden, sodass Prompts zumindest so sichtbar sind, dass sie den Nutzern auffallen, aber nicht so übersteigert gestaltet sind, dass sie den Nutzer von seiner eigentlichen Tätigkeit ablenken. Die in dieser Arbeit eingesetzten Prompts sind im Sinne der Sichtbarkeit eher als wenig sichtbar einzuordnen, was die Interviews bestätigen, in denen Nutzer sagten, dass sie die Funktionen zur Unterstützung von Reflexion nur in wenigen Fällen wahrgenommen haben (siehe Abschnitt 4.4.5).

#### **6.1.9.2 Startseite der Community-Plattform**

Die Startseite der Community of Practice Plattform zeigt eine Übersicht der neuesten Beiträge und zeigt Beiträge, die in der letzten Zeit viele Bewertungen (bspw. für wertvolle Beiträge) bekommen haben. Ebenso wurden auf der Startseite der Plattform Prompts angezeigt (siehe Abbildung 4-3), um Nutzer dazu zu motivieren, direkt von dort aus neue Themen zu erstellen.

Die Prompts haben auf der Startseite nie zu einem neuen Thema geführt bzw. sind in Zusammenhang mit einem neuen Thema aufgetaucht. Dies kann zwei Gründe haben: Zum einen war die Formulierung der Prompts auf der Startseite auf neue Erlebnisse oder überraschende Erlebnisse fokussiert. Diese Prompts haben jedoch im Bereich des Forums nicht funktioniert und können dementsprechend keinen Einfluss gehabt haben. Zum anderen passt vermutlich die Startseite als Ort neue Beiträge zu verfassen nicht ins mentale Modell der Nutzer, sodass diese lieber über die Ansicht des Forums neue Beiträge einstellen. Dies kann dadurch verstärkt sein, dass der Prompt visuell von der Übersicht neuer Beiträge zu stark entfernt war und eine dichtere räumliche Nähe sinnvoller ist.

Obwohl Prompts nicht zu einer erhöhten Menge an Beiträgen geführt haben (siehe Abschnitt 6.1.1), ist es durchaus sinnvoll, auf der Startseite einer Community of Practice Plattform einen Einstieg in die Diskussion zu erleichtern. Die Vermutung ist, dass das Konzept verfeinert werden muss, um einen Erfolg zu erzielen. Im Abschnitt 7.2 wurde ein Vorschlag dafür erarbeitet.

## **6.2 Diskussion der Ergebnisse der Studien hinsichtlich der Forschungsfrage**

An dieser Stelle soll eine Übersicht der Ergebnisse hinsichtlich der Forschungsfrage, inwieweit Prompts unterschiedliche Aspekte von Reflexion unterstützen können, aufgestellt werden.

Hinsichtlich der im Konzept aufgestellten Ziele für die Unterstützung von Reflexion ist festzustellen, dass die Daten aus der Feldstudie (siehe Kapitel 4) in den meisten Fällen nicht ausreichen, um konkrete Aussagen treffen zu können. Die Ziele für das Beitragen eigener Erfahrungen (Z6) und dem Erstellen von Lösungsvorschlägen (Z8) haben insbesondere in den Crowdsourcing-Studien vielversprechende Ergebnisse gezeigt (siehe Abschnitte 5.1, 5.2 und 5.3). Nutzergenerierte Prompts scheinen ebenfalls erfahrungsbasierte Lösungsvorschläge begünstigen zu können (siehe Abschnitt 6.1.2). Für das Ziel, das eigene Lernen zu dokumentieren bzw. lernen anzuregen (Z9 und Z10) haben die beiden Anwendungsfälle in der Feldstudie unterschiedliche Ergebnisse gezeigt. Daraus lässt sich nicht zweifelsfrei ableiten, dass Prompts dieses Ziel unterstützen können. Eine Tendenz ist vorhanden und weitere Studien sind notwendig, um bspw. andere Formulierungen dieser Prompts zu testen. Alle übrigen Ziele lassen sich aufgrund der niedrigen Datenmenge nicht zweifelsfrei als förderbar oder explizit nicht-förderbar beschreiben. Hierzu sind weitere Studien notwendig. Aufgrund der Vielzahl unterschiedlicher Ziele ist entweder eine große Menge Probanden in einer Feldstudie notwendig oder mehrere kleinere fokussierte Crowdsourcing-Studien.

Die Ergebnisse der Studien haben allerdings weitere Dimensionen der Wirkung von Prompts für kollaborative Reflexion aufgezeigt, die in den vorherigen Abschnitten bereits diskutiert wurden. Es zeigt sich, dass nutzergenerierte Prompts eine Möglichkeit sind, in manchen Anwendungsfällen die Antwortrate der Probanden zu erhöhen, und häufig zusammen mit erfahrungsbasierten Lösungsvorschlägen auftauchen (siehe Abschnitt 6.1.2).

Zusätzlich zeigen die Studien, dass Prompts für Lösungsvorschläge bereits von Anfang an gezeigt werden können und signifikant bessere Lösungsvorschläge hervorrufen als später in einem Thema (siehe Abschnitt 6.1.4). Der Unterschied in der Qualität ist

allerdings sehr gering. Die entstandenen Lösungsvorschläge in beiden Teilstudien sind bei den in dieser Arbeit genutzten Prompts und in den abgebildeten Situationen nicht in der Lage viele Ideen hervorzurufen, die als *neu* bewertet werden (siehe Abschnitt 6.1.5).

Wenn Prompts nach Unerwartetem fragen, was einer der Auslöser in der Theorie von Reflexion ist, dann wird dies von Nutzern nicht dazu genutzt neue Beiträge zu verfassen (siehe Abschnitt 6.1.6). In der genutzten Community of Practice Plattform haben diese Prompts keine messbare Wirkung gezeigt. Aufgrund der geringen Datenmenge sollte dieser Fund in weiteren Studien bestätigt werden.

Die Anzahl der Wörter war häufig ein Faktor in den Studien (siehe Abschnitt 6.1.7). Obwohl dies nicht ein aus der Reflexionstheorie abgeleiteter Faktor ist, sondern mehr eine Art Hygienefaktor für qualitative Beiträge darstellt, lässt sich diese Erkenntnis in weiteren Prompts nutzen (siehe 7.2).

### **6.3 Einschränkungen der Arbeit**

Spezifische Einschränkungen der einzelnen Studien wurden bereits in den Abschnitten 4.7 (Studie 1), 5.1.7 (Teilstudie 2.1), 5.2.7 (Teilstudie 2.2) und 5.3.7 (Teilstudie 2.3) diskutiert. An dieser Stelle werden zusätzlich allgemeine Einschränkungen für die Ergebnisse bzw. Schlussfolgerungen dieser Arbeit beschrieben.

Die Nutzer aus Studie 1 arbeiten in öffentlichen Verwaltungen und befinden sich demnach nicht mehr in einer Ausbildung oder einem Studium. Bei den Nutzern aus der Crowdsourcing-Studie ist die Information nicht bekannt. Diese können weiterhin Studenten oder reguläre Arbeitnehmer sein. Studien haben gezeigt, dass die Ergebnisse aus Studien mit Studenten nicht zwingend übertragbar sind auf Erwachsene im Berufsleben (Ward 1993). Aus diesem Grund sollten die Ergebnisse aus den Crowdsourcing-Studien in einem praktischen Umfeld repliziert und bestätigt werden.

Reflexion ist eine Form informellen Lernens. Bei den durchgeführten Studien gibt es allerdings keine Möglichkeit, eine Auswirkung des Diskutierten auf die tägliche Praxis oder das Gelernte zu ziehen. Die Daten zeigen lediglich, dass die Prompts zu bestimmten Eigenschaften in den Texten geführt haben. Inwieweit daraus Personen in den beiden Anwendungsfällen in Studie 1 (siehe Kapitel 4) gelernt haben und inwieweit das Gelernte einen Einfluss auf ihre tägliche Praxis hat, lässt sich anhand der Daten nicht beurteilen. Hinsichtlich der Studien, die Prompts in der Crowdsourcing-Umgebung evaluiert haben (siehe Abschnitt 5.2), lässt sich dies ebenfalls nicht beantworten. Da es sich um ein fiktives, aber realistisches Thema handelt, gibt es in diesem Szenario keine einzelne

Person, die dieses Problem hatte, und die anschließend befragt werden könnte, inwieweit lernen stattgefunden hat. Dementsprechend zeigen die Ergebnisse eher inwieweit sich die Aspekte der Reflexion unterstützen lassen. Zu welchem Grad diese Faktoren zum Lernen beitragen lässt sich jedoch nicht beantworten.

Zusätzlich können manche Unterschiede in den Bewertungen daher stammen, dass Personen unterschiedlich lernen und nicht jede Person hauptsächlich reflexiv lernt (Kolb 1984). Dementsprechend können unterschiedliche Personen unterschiedlich stark von durch Prompts unterstützter schriftlicher kollaborativer Reflexion profitieren.



## **7 Empfehlungen zur Implementierung von Prompting als Unterstützung von Reflexion**

In diesem Kapitel werden Empfehlungen für den Einsatz von Prompts aus den Ergebnissen dieser Arbeit abgeleitet. Weiterhin gibt es mehrere Ideen, die sich aus den Ergebnissen ableiten, welche zunächst noch in weiteren Studien geprüft werden müssen. Diese werden in einem weiteren Abschnitt skizziert.

### **7.1 Empfehlungen auf Basis konkreter Ergebnisse**

Die dargestellten Empfehlungen lehnen sich an die Ergebnisse aus der allgemeinen Diskussion in Kapitel 5.4 an.

Zunächst ist zu unterstreichen, dass die Sichtbarkeit der Prompts sehr wichtig ist (Butler und Winne 1995). Dieser Faktor war in der Crowdsourcing-Studie weniger im Fokus, weil dort der Prompt direkt beim Verfassen einer Antwort sichtbar ist. Das Ziel bei der Positionierung der Prompts war es, die Prompts dort zu zeigen, wo der Nutzer eine Aktivität durchführt (McKenzie-Mohr 2000), sprich in der Nähe der Stelle, wo der Nutzer einen Beitrag verfasst. Dementsprechend war der Fokus der Arbeit mehr auf der Unterstützung des Verfassens von Beiträgen und weniger auf dem Anregen von neuen Beiträgen.

Studie 1 hat gezeigt, dass die Nutzer Prompts in der dargestellten Form in einer Community of Practice Plattform in öffentlichen Verwaltungen zur Unterstützung von Reflexion annehmen. Obwohl die Anzahl der Beiträge eher gering ist, was die Aussagekraft der Ergebnisse schwächt, kann die genutzte Form der Unterstützung empfohlen werden. Somit können die Nutzer beim Verfassen neuer Beiträge unterstützt werden. Die Anzahl der Beiträge wird durch Prompts eher nicht erhöht.

Hinsichtlich konkreter Prompts haben sich sowohl der Lösungs-Prompt „*Have you ever been in a similar situation? Enter a comment that describes what you did.*“ als auch der Erfahrungs-Prompt „*Enter a comment. Describe a solution to the problem above and why you think this is a solution.*“ aus Teilstudie 2.1 (siehe Abschnitt 5.2) bewährt. Beide Prompts haben in der Crowdsourcing-Studie die entsprechende Wirkung gehabt. Obwohl die Lösungs-Prompts besser bewertete Lösungsvorschläge als die Erfahrungs-Prompts hervorbringen, sind die Lösungsvorschläge der Erfahrungs-Prompts nicht per se „schlecht“ bewertet. Die Unterschiede in der Bewertung waren sehr gering. Dementsprechend stellen sie keinen Nachteil dar und der Prompt kann gemäß den

Ergebnissen dieser Arbeit problemlos verwendet werden, um Lösungsvorschläge in einer Diskussionsumgebung, wie einer Community of Practice, anzuregen.

Die Ergebnisse zeigen auch, dass Lösungs-Prompts bereits von Anfang an eingesetzt werden können und teils besser bewertete Lösungsvorschläge erzeugen als zu einem späteren Zeitpunkt, zu dem eine Diskussion bereits läuft. Die Bewertung der Lösungsvorschläge, die in einem neuen Thema ohne Diskussionsbeiträge gegeben wurde, ist nur minimal aber signifikant besser als die Texte, die zu einem späteren Zeitpunkt, zu dem Diskussionsbeiträge vorhanden sind, entstanden sind. Dementsprechend ist die Empfehlung, den Lösungs-Prompt direkt von Anfang an einzusetzen und dann weiter in laufenden Diskussionen beizubehalten.

Prompts, die nach unerwarteten oder überraschenden Erlebnissen in der jüngeren Vergangenheit fragen, um so eine Reflexion auszulösen, haben in Studie 1 nicht funktioniert. Dementsprechend wird empfohlen, diese Art Prompts nicht einzusetzen.

Die Wirkung der Prompts wurde nur in einer Crowdsourcing-Studie bestätigt. Da in der Studie ein Diskussionsbereich simuliert wurde, kann angenommen werden, dass diese Prompts ihre Wirkung auch in anderen Systemen wie bspw. in einer Community of Practice Plattform, haben können. Dies muss in weiteren Studien bestätigt werden.

Zusätzlich hat der Studienzeitraum von drei Monaten in Studie 1 gezeigt, dass dort noch keine Effekte von Over-Prompting (Nückles, Hübner und Renkl 2006) zu beobachten sind. Dementsprechend kann dieser Zeitraum für diese Art Unterstützung beibehalten werden. Bei längeren Zeiträumen empfiehlt es sich, die Prompts nicht bei jedem Seitenaufruf anzuzeigen.

Nutzergenerierte Prompts hatten teilweise eine höhere Antwortrate als die systemgenerierten Prompts. Dementsprechend ist die Empfehlung, diese Möglichkeit für Nutzer in Forenumgebungen beizubehalten. Im kommenden Abschnitt 7.2 wird das Konzept der nutzergenerierten Prompts grafisch erweitert, um möglicherweise den Effekt der Prompts zu erhöhen.

## **7.2 Vorschläge für Prompt-Design**

Neben den direkten Erkenntnissen aus dieser Arbeit gibt es noch mehrere andere interessante Aspekte, die in künftigen Studien untersucht zu werden sollten. Diese wurden teilweise bereits in der Diskussion in Kapitel 5.4 grob beschrieben, und werden weiter detailliert.



## 7.2.1 Vorschlag 1: Hervorhebung von nutzergenerierten Prompts

Ein interessanter Aspekt aus Studie 1 ist, dass in einem Anwendungsfall die nutzergenerierten Prompts zusammen mit höheren Antwortraten auftraten und häufig mit erfahrungsbasierten Lösungsvorschlägen auftauchten (siehe Abschnitt 6.1.2).

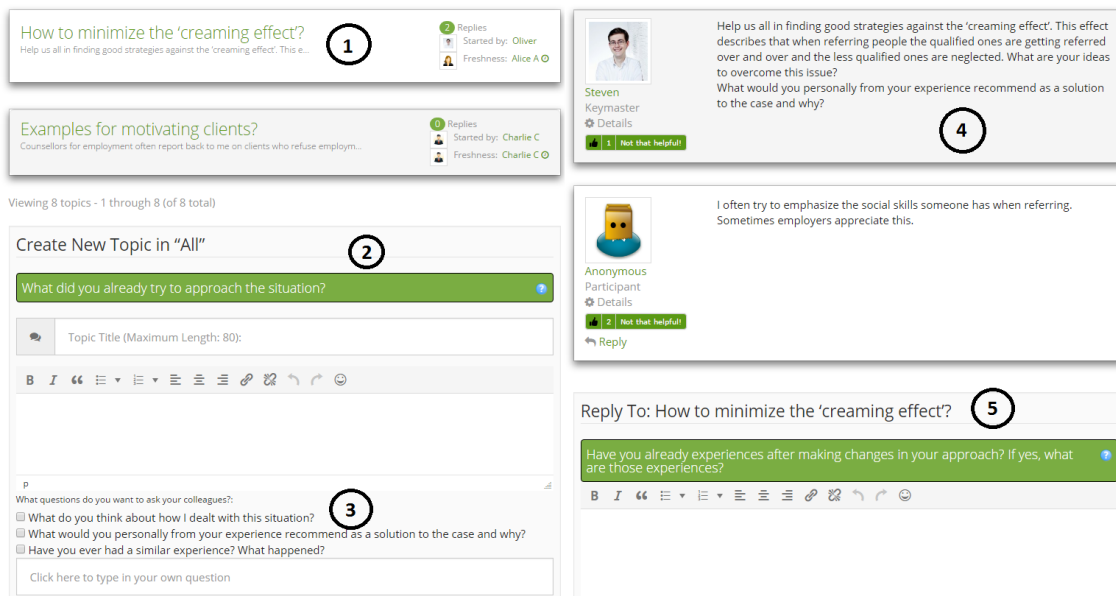


Abbildung 7-1: Prompts im Bereich der Foren (grün hinterlegt; ursprüngliches Design)

Bisher werden nutzergenerierte Prompts an das Ende eines Beitrags gehängt und visuell hervorgehoben. Sie ersetzen dann den systemgenerierten Prompts (siehe Abbildung 7-1; (3) Auswahl einer Frage; (4) Frage normal im Text angehängt und (5) Frage als nutzergenerierter Prompt). Von außen ist aktuell schwierig erkennbar, dass es sich um einen nutzergenerierten Prompt handelt. Eine Idee ist, diese Sichtbarkeit zu erhöhen und deutlicher hervorzuheben, dass der angezeigte Prompt nicht vom System generiert ist, sondern von einem echten Nutzer stammt. In Abbildung 7-2 ist ein erster Entwurf dafür abgebildet und zeigt die Perspektive eines Nutzers, der ein Thema mit einem nutzergenerierten Prompt liest. Es wurde ein nutzergenerierter Prompt ausgewählt, der im Text sichtbar ist und unten beim Texteditor visuell hervorgehoben wird. Zusätzlich zeigt das System den Namen und den Avatar des Autors an. Das System prüft den Text des angehängten nutzergenerierten Prompts, sodass bei Fragen ein „<Name> asks“ vor den Prompt eingefügt und bei Aussagen ein „<Name> says“ vor dem Prompt eingeblendet wird. Auf diese Weise sollen die Fragen natürlicher wirken und sich von systemgenerierten Prompts abheben. Weitere Studien müssen zeigen, ob diese Art der Darstellung die Effekte nutzergenerierter Prompts erhöht bzw. verstärkt.

Issue with a client

The image shows a user interface for a client issue. At the top, there is a profile card for 'Oliver', a 'Keymaster', with a 'Details' link and a timestamp of '12.03.2018 at 10:51'. Below this is a green button that says '0 This is useful!'. The main content area is titled 'Issue with a client' and contains a green bar with the text 'Oliver asks "What would you personally from your experience recommend as a solution to the case and why?"'. Below this bar is a rich text editor with various formatting icons (bold, italic, quote, list, link, etc.) and an 'Upload File' button. The editor is currently empty. Below the editor is a 'Tags:' section with a text input field. At the bottom, there are two checkboxes: 'Notify me of follow-up replies via email' (checked) and 'Post Anonymously' (unchecked). A yellow 'Submit' button is located at the bottom right of the form.

**Abbildung 7-2: Prototyp von verbesserten nutzergenerierten Prompts  
(eigener Entwurf; neues Design)**

### 7.2.2 Vorschlag 2: Positionierung von Prompts

Die Positionierung der Prompts wurde kritisch diskutiert (siehe Abschnitt 6.1.9). Die Prompts in Studie 1 und in Teilstudie 2.2 waren in der Nähe des Eingabefelds für neue Beiträge dargestellt. Dies setzt einen Schwerpunkt auf die Unterstützung bei dem Verfassen neuer Beiträge, aber weniger beim Anregen neuer Beiträge, da der Prompt so oft nicht sofort beim Öffnen eines Seitenabschnitts im Sichtbereich des Nutzers ist. Um das Anregen neuer Beiträge zu unterstützen, empfiehlt es sich, den Prompt so anzuzeigen, dass er auf der Plattform direkt beim Öffnen eines neuen Abschnitts im Sichtfeld ist.

Innerhalb eines Themas kann dies bspw. unterhalb der ersten Antwort sein. Die Idee ist, den Prompt so zwei Mal innerhalb eines Themas anzuzeigen: direkt im Sichtfeld, um einen neuen Beitrag anzuregen und in der Nähe des Eingabefelds für Beiträge. Abbildung 7-3 zeigt denselben Prompt sowohl zu Beginn eines Themas als auch am Eingabefeld. Zusätzlich werden die Prompts mit einem Icon hervorgehoben. Die Hypothese ist, dass bei einer zusätzlichen Darstellung der Prompts direkt im Sichtfeld des Nutzers mehr

Not willing to take job – motivation?

### Abbildung 7-3: Entwurf einer Prompt-Wiederholung (eigener Entwurf)

301

springen. Dies könnte die Navigation erleichtern, wenn sehr viele Themen vorhanden sind. Die Hypothese ist, dass dadurch Nutzer leichter auf den Gedanken gebracht werden können einen neuen Beitrag zu verfassen. Dies müsste ebenfalls in weiteren Studien überprüft werden. Hier ist allerdings zu klären, welchen Inhalt die Prompts haben, da Fragen hinsichtlich unerwarteter Begebenheiten in der Arbeit als Prompt-Inhalte in Studie 1 nicht funktioniert haben (siehe Abschnitt 6.1.6).

Eine andere Möglichkeit ist, das Eingabefeld für neue Themen direkt am Anfang einer Seite anzuzeigen. Hier sollte dieses Feld minimiert dargestellt werden, da die meisten Seitenaufrufe lesender Art sind und sich die Nutzer zunächst die Liste der Beiträge anschauen, um neue Themen zu lesen oder zu kommentieren. Wenn das Eingabefeld oberhalb der Liste der Beiträge ist, muss der Nutzer jedes Mal über dieses Feld hinwegnavigieren. Aus diesem Grund sollte eine Balance zwischen einem leichten Zugang zum Textfeld und einem schnellen Navigieren zu den Beiträgen gefunden werden.

### **7.2.3 Vorschlag 3: Prompts zur Vermeidung sehr kurzer Beiträge**

Die Ergebnisse aus Teilstudie 2.1 (siehe Abschnitt 5.1), 2.2 (Abschnitt 5.2) und 2.3 (Abschnitt 5.3) zeigen, dass die Anzahl der Wörter ein Faktor hinsichtlich der Codes für Erfahrungen oder Lösungsvorschläge ist und einen Einfluss bei der Bewertung der Qualität der Lösungsvorschläge hat (siehe Abschnitt 6.1.7). Aus diesen Ergebnissen lässt sich ableiten, dass Beiträge mit sehr wenigen Wörtern Schwierigkeiten haben entweder bestimmte Aspekte, wie Erfahrungen oder Lösungsvorschläge, zu transportieren oder auch qualitative Lösungsvorschläge beizutragen. Die Hypothese ist, dass Prompts Nutzer dazu bewegen können, einem Beitrag mehr Details hinzuzufügen.

In Abbildung 7-4 wird ein möglicher Prompt dargestellt, der den Nutzer darüber informiert, dass der Beitrag zu kurz ist, und es sinnvoll sein könnte mehr Wörter hinzuzufügen. Rein textuell informiert der Prompt über Ergebnisse aus einer Studie und regt zu mehr Details an. In einer Studie wurde anstatt der Referenz auf Forschungsergebnisse eine ärztliche Empfehlung erfolgreich genutzt, um körperliche Aktivitäten anzuregen (Kaczynski, Wilhelm Stanis und Hipp 2014). Eine andere Idee für einen Prompt ist das Auflisten möglicher Vorteile eines umfangreicheren Beitrags, wie bspw. das Hinzufügen wichtiger Aspekte qualitativer Lösungen (vgl. Abschnitt 5.3.2.4). Rein technisch kann das System, sobald ein Beitrag abgeschickt werden soll, prüfen, wie viele Wörter dieser enthält und ab bestimmten Schwellwerten unterschiedliche Prompts anzeigen. Als erste Annäherung an mögliche Schwellwerte könnten die

Wörtergruppierungen aus dieser Arbeit genutzt werden. Dabei ist allerdings zu beachten, dass diese Gruppierungen spezifisch für diese Studien und die Crowdsourcing-Umgebung sind. Je nach Thema variiert die Länge eines Beitrags, weshalb es auch sinnvoll sein kann eine dynamische Einteilung der Wörtergruppierungen vorzunehmen. In dieser Abbildung ist der Prompt sehr allgemein. Falls bspw. nur solche Prompts angezeigt werden sollen, sollten Lösungsvorschläge eingebracht werden, müsste das System allerdings den Inhalt des Beitrags dazu kennen. In der vorliegenden Arbeit wurden textuelle Inhalte ausschließlich manuell bewertet, was nur post-hoc möglich ist und sehr zeitaufwendig ist. Es gibt unterschiedliche Arbeiten, die eine automatisierte Analyse von Texten hinsichtlich Reflexion (Ullmann 2015a) oder anderen Kategorien ermöglichen (Pennebaker u. a. 2015; Fast, Chen und Bernstein 2016). Diese sind im Vergleich zu einem dedizierten Kodierschema für Reflexion, wie das eingesetzte Schema, zu ungenau, um Reflexion zu analysieren (vgl. auch Kapitel 3).

Abbildung 7-4: Entwurf eines Hinweises für die Anzahl der Wörter (eigener Entwurf)

#### 7.2.4 Vorschlag 4: Konvergenzphase zur Verarbeitung der Lösungsvorschläge

Die Ergebnisse haben gezeigt, dass es möglich ist, direkt zu Beginn eines neuen Themas, bereits bevor Antworten geschrieben bzw. Erfahrungen ausgetauscht wurden, nach Lösungsvorschlägen zu fragen (siehe Abschnitt 6.1.4). Dementsprechend können Lösungs-Prompt permanent angezeigt werden und haben eine hohe Wahrscheinlichkeit, dass diese neue Lösungsvorschläge anregen können (siehe Abschnitt 6.1.3). Eine Vermutung ist, dass dies nicht über die komplette Dauer einer Diskussion sinnvoll ist und man ähnlich wie bei einem Brainstorming eine Konvergenz-Phase benötigt, bei der die gesammelten Ideen zusammengeführt bzw. selektiert werden (Santanen 2005). Das längere Anzeigen des Lösungs-Prompts stellt dann eine Divergenz-Phase dar, in der mehrere Lösungsvorschläge gesammelt werden. Bei (kollaborativer) Reflexion wird

ebenfalls eine Idee selektiert und aufbereitet, sodass diese danach für den Einsatz in der Praxis vorbereitet werden kann. Wenn der Lösungs-Prompt über die ganze Dauer der Diskussion hinweg angezeigt wird, kann es passieren, dass dieser Schritt nicht in der benötigten Tiefe behandelt wird. Hier sind weitere Studien notwendig, um diesen Fall zu evaluieren.

## 8 Zusammenfassung

In diesem Kapitel wird zunächst die vorliegende Arbeit reflektiert. Abschließend erfolgt ein Fazit der durchgeführten Arbeit.

### 8.1 Möglichkeiten für weitere Forschung

Aus den Ergebnissen der vorliegenden Arbeit ergeben sich unterschiedliche Anschlussmöglichkeiten für weitere Studien, welche hier im Folgenden kurz skizziert werden.

Die genutzten Prompts in der vorliegenden Arbeit sind alle systemgeneriert und werden nach einer groben Heuristik an Nutzer ausgeliefert. Dabei liegt für jedes der im Konzept definierten Ziele genau ein Prompt vor. Aufgrund der Anzahl der Ziele und der limitierten Anzahl der genutzten Prompts, kann es hier relativ bald zu Ermüdungserscheinungen der Nutzer hinsichtlich der Beachtung der Prompts kommen (Kocielnik u. a. 2018). Als weitere Forschungsrichtung bietet sich die Personalisierung von Prompts an.

Bei nutzergenerierten Prompts hat der Nutzer in einer Community of Practice die Möglichkeit, einen Prompt für die Leser seines geschriebenen Themas zu gestalten. Dieser weicht dann von systemgenerierten Prompts ab und kann wesentlich besser auf den jeweiligen Kontext oder auch die Zielrichtung des Beitrags passen. Eine erste Skizze, wie diese in Zukunft aussehen können, findet sich in Abschnitt 7.2.1. Hier wird eine fokussierte Studie benötigt, um zu evaluieren, welche Effekte bei Nutzung von nutzergenerierten Prompts vorliegen. Die Studie sollte dabei direkt nutzergenerierte und systemgenerierte Prompts kontrastieren.

Eine andere Form der Personalisierung kann auch eine gezieltere Ansprache der Nutzer sein, was ebenfalls eine Überlegung bei der Gestaltung von Dialogen für Hilfestellungen als Anregung für physische Aktivitäten (Kocielnik u. a. 2018) oder bei der Ansprache in Nudges bei Online-Kursen ist (Dimitrova u. a. 2017). Dies kann auch eine Möglichkeit sein, den Nutzer in Prompts anzusprechen und womöglich eher dazu führen, dass der Nutzer die Inhalte des Textes in seine Handlung aufnimmt.

Da in Studie 2 die Anzahl der Wörter mehrfach einen Einfluss hatte, lohnt es sich zu evaluieren, inwieweit Prompts für die Anzahl von Wörtern eine zusätzliche aber indirekte Hilfestellung bei (kollaborativer) Reflexion darstellen können (Designvorschlag in Abschnitt 7.2.3). Das Ziel dieser Prompts wäre Nutzer darauf aufmerksam zu machen, dass bspw. ein Beitrag zu kurz ist und dadurch womöglich bestimmte Details, wie bspw. Lösungsvorschläge, fehlen. Auf der anderen Seite können Prompts in der Studie ebenfalls

auch testen, ob es Sinn macht, Prompts für zu umfangreiche Beiträge zu zeigen, da solche Beiträge seltener gelesen werden (Wise, Zhao und Hausknecht 2013). In einer Studie könnten Prompts für die Länge der Beiträge zusätzlich zu den übrigen Prompts, wie bspw. für Lösungsvorschläge, genutzt werden.

Aus den Ergebnissen aus Studie 1 ist wenig bekannt, inwieweit Over-Prompting einen Effekt gehabt haben kann. Wenn kollaborative Reflexion längerfristig unterstützt werden soll, um bspw. mangelnde Motivation zu minimieren (Clarà 2015) oder Personen in der Praxis bei der Durchführung zu unterstützen (Platzer, Snelling und Blake 1997), müssen die Prompts auch langfristig funktionieren. Da die Wirkung von Prompts über die Zeit abnehmen kann (van Oldenbeek u. a. 2019), ist es wichtig, zu wissen, in welchem Zeitrahmen, welche Minderung der Effekte eintritt. Hierzu sind ebenfalls weitere Studie notwendig.

## **8.2 Fazit**

Die vorliegende Arbeit behandelt die Unterstützungsmöglichkeiten kollaborativer Reflexion durch Prompts. Kollaborative Reflexion wird als Form informellen Lernens aus der eigenen Erfahrung verstanden, bei der eine Analyse der Erfahrung durch eine erlebte Diskrepanz des Erlebten zum Erwarteten ausgelöst wird (siehe Abschnitte 2.4 und 2.5). Die Reflexion ist zielgerichtet und strebt die Entwicklung neuer Hypothesen an, wie in künftigen ähnlichen Situationen gehandelt werden kann. Diese Form informellen Lernens ist an Arbeitsplätzen stark verbreitet. In der Praxis wird häufig eine Community of Practice als Ort für Reflexion genutzt (siehe Abschnitt 2.1). Darunter versteht man Gruppen, die eine gemeinsame Praxis haben (bspw. die gleiche Spezialisierung in einem Beruf) und deren Ziel das gemeinsame Lernen ist. Gruppen schließen sich oft in solchen Communities of Practice zusammen, um Erfahrungen zu teilen oder Herausforderungen zu diskutieren. Dies ist ein Ort, der für informelles Lernen genutzt wird. In der Praxis unterliegt (kollaborative) Reflexion aber einigen Herausforderungen. Es fehlt dort oft die Zeit, um reflektieren zu können, bzw. häufig wird eine inhaltliche Unterstützung benötigt, da der Prozess der Reflexion in der Praxis schwierig systematisch durchzuführen ist (siehe Abschnitt 2.8). Dies lässt sich auch an den Ergebnissen in Studie 1 bei den Anwendungsfällen in Slowenien ablesen, da dort auch zeitliche Restriktionen als Grund für fehlende Reflexion auf der Plattform genannt wurde (siehe Abschnitt 4.4.5). Eine skalierbare und personalisierbare Form der Unterstützung sind Prompts, welche in der vorliegenden Arbeit durch textuelle Nachrichten verkörpert werden, die Nutzern angezeigt werden, wenn Nutzer schriftliche Beiträge verfassen. Da Prompts bisher häufiger zur Unterstützung von Reflexion im Bildungsbereich und weniger in der



beruflichen Praxis eingesetzt werden, stellt sich die Frage, inwieweit sich Prompts zur Unterstützung von Reflexion bzw. von Aspekten kollaborativer Reflexion (bspw. Erfahrungsaustausch) in einer Community of Practice eignen.

Zur Beantwortung dieser Fragestellung wurden unterschiedliche Studien durchgeführt. Aus diesen Studien wurden unterschiedliche Empfehlungen zum Design von Prompts zur Unterstützung von kollaborativer Reflexion abgeleitet (siehe Abschnitt 7.1). Es wurde eine Studie in einer Community of Practice Plattform inklusive Prompts als Mechanismus zur Unterstützung von kollaborativer Reflexion durchgeführt (siehe Kapitel 4). In dieser Studie sind zwei Anwendungsfälle vorhanden. Die Ergebnisse zeigen, dass Prompts in dieser Situation nicht zu mehr neuen Themen geführt haben und sich somit in der genutzten Form nicht dazu eignen, die Beteiligung in Form von neuen Themen zu erhöhen. Wenn Nutzer selbst Fragen formulieren und an die eigenen Themen anhängen, die dann für die Antwortenden gezeigt werden, scheint eine höhere Antwortrate vorzuliegen. Diese Form der Prompts wird in der vorliegenden Arbeit als nutzergenerierte Prompts beschrieben. Diese Form der Prompts haben zusätzlich einen höheren Zusammenhang zu erfahrungsbasierten Lösungsvorschlägen. Dementsprechend scheint dies eine relevante Form von Prompts als Unterstützungsmöglichkeit für kollaborative Reflexion zu sein. Weitere Studien sind notwendig, um mehr Klarheit zu erzielen. Allgemein zeigen die beiden Anwendungsfälle, dass Prompts bei 40-46 % der Beiträge einen inhaltlichen Zusammenhang zu den Texten aufweisen, was impliziert, dass ein Effekt der Prompts vorhanden ist. Aufgrund der geringen Datenmenge ist keine Aussage möglich, welcher Prompt in der Praxisumgebung, zu welchen Effekten in den Texten geführt hat. Einzig ein Prompt, der Lernen anregen sollte, trat in einem Anwendungsfall laut der Inhaltsanalyse mit Texten auf, in denen Lernen abgebildet war. In dem anderen Anwendungsfall wurde dieser Effekt nicht beobachtet. Dies kann allerdings ebenfalls der geringen Datenmenge geschuldet sein. Weitere Studien sind notwendig, um einen möglichen Effekt zu identifizieren.

Um eine detailliertere Untersuchung der Effekte von Prompts durchzuführen, wurden zwei Crowdsourcing-Studien mit unterschiedlichen Prompts durchgeführt (siehe Abschnitte 5.1 und 5.2). Die erste Studie hat Prompts für Lösungen und Erfahrungen miteinander verglichen. Die Ergebnisse der Studie zeigen, dass beide Prompts die entsprechenden Aspekte von Reflexion anregen. Diese Prompts unterscheiden sich inhaltlich sehr stark. In einer Crowdsourcing-Studie wurden Prompts für Lösungswissen, die inhaltlich stärker zusammenliegen, evaluiert. Hier wurde kein signifikanter Unterschied hinsichtlich der Wirkung der Prompts identifiziert. Dementsprechend sind Nutzer in der Lage, die groben Konzepte der Prompts gut zu erfassen und umzusetzen. Feinere Details in den Prompts finden sich nicht immer innerhalb der Texte wieder.

Es wurde bei den Crowdsourcing-Studien ebenfalls verglichen, zu welchem Zeitpunkt in einer Diskussion Prompts für Lösungsvorschläge genutzt werden können. Es wurden zwei Konditionen verglichen: ein neues Thema ohne Antworten und ein Thema mit drei vorhandenen Antworten. Die Ergebnisse zeigen, dass sich die Menge der Lösungsvorschläge innerhalb der durch Nutzer verfassten Texte zwischen den beiden Konditionen nicht unterscheidet. Um gleichzeitig die Qualität der Lösungsvorschläge zu untersuchen, wurde eine weitere Studie durchgeführt, in welcher die Lösungsvorschläge systematisch bewertet wurden (siehe Abschnitt 5.3). Die Ergebnisse zeigen, dass die Lösungsvorschläge in einem neuen Thema ohne Antworten signifikant, wenn auch mit kleinem Unterschied, besser bewertet werden. Dies bedeutet, dass Prompts als Form der Unterstützung für kollaborative Reflexion direkt von Anfang an sinnvoll sind. In der Literatur wurde bislang ein Erfahrungsaustausch als elementarer Bestandteil von kollaborativer Reflexion genannt (siehe Abschnitt 2.5). Die Ergebnisse zeigen, dass dieser zwar wichtig sein mag, aber nicht der Möglichkeit, direkt qualitative Lösungsvorschläge zu erzeugen, im Weg steht.

Die Literatur von Reflexion beruht darauf, dass Reflexion ausgelöst wird, indem eine Diskrepanz zwischen Erwartetem und Erlebtem existiert, welches dann oft zu Verwunderung oder einer Überraschung führt (siehe Abschnitt 2.4). Eine Idee in der vorliegenden Arbeit war es, diesen Auslöser als Gegenstand von Prompts zu nutzen, um in der Praxis Reflexion zu fördern. Innerhalb der Studie in der Praxisumgebung (siehe Kapitel 4) trat allerdings keiner der dazu passenden Prompts zusammen mit einem Text auf. Dies bedeutet, dass in der Theorie das Überraschende zwar ein Auslöser für Reflexion sein mag, dies aber für die Nutzer kein Anhaltspunkt war, der zum Verfassen eines neuen Beitrags führte. Es ist möglich, dass die Empfänger außerhalb der Community of Practice Plattform über die Prompts nachgedacht oder mit Kollegen darüber gesprochen haben. Dazu konnten aber keinerlei Daten gesammelt werden. Hier sind weitere Studien notwendig, um zu ermitteln, welche Botschaft in Prompts ein passender Auslöser sind, um die Empfänger dazu zu motivieren, einen neuen Beitrag in einer online Community of Practice zu verfassen.

In den Crowdsourcing-Studien stand die Anzahl der Wörter innerhalb eines Textes oft in Zusammenhang mit bestimmten Inhalten von Texten, wie Erfahrungen oder Lösungsvorschlägen (siehe Abschnitte 5.1 und 5.2). Zusätzlich wurden mehrere Zusammenhänge zwischen der Anzahl der Wörter und den Qualitätsmerkmalen der Lösungsvorschläge identifiziert (siehe Abschnitt 5.3). In allen Studien trat das gleiche Muster auf: Beiträge mit sehr wenigen Wörtern (0-40) haben die geringste Wahrscheinlichkeit, dass bestimmte Inhalte wie Erfahrungen oder qualitative Lösungsvorschläge vorhanden sind. Die Wahrscheinlichkeit dafür steigt, wenn die

Anzahl der Wörter steigt. Ab einer Anzahl von 81 Wörtern nahm dieser Effekt ab. Dies zeigt deutlich, dass eine minimale Anzahl an Wörtern notwendig ist, um bestimmte Inhalte zu transportieren, aber dass es nach oben hin keine lineare Skalierung gibt. Dies könnte man in weiteren Prompts umsetzen, die bspw. Nutzer dazu animieren, bestimmte Aspekte in ihren Texten abzudecken, um die Wahrscheinlichkeit auf einen qualitativen Lösungsvorschlag zu erhöhen.

Die Ergebnisse der vorliegenden Arbeit zeigen, dass sich bestimmte Aspekte kollaborativer Reflexion, wie Erfahrungen und Lösungsvorschläge, in einer Crowdsourcing-Umgebung durch Prompts sehr gut anregen lassen.

Es werden zusätzlich unterschiedliche Richtungen für künftigen Arbeiten aufgezeigt (siehe Abschnitt 7.2): Zum einen wirkte das Konzept der nutzergenerierten Prompts in der Praxisstudie in der Community of Practice Plattform vielversprechend und sollte weiter evaluiert werden. Zum anderen scheint Unerwartetes zwar ein Auslöser für Reflexion zu sein, aber nichts, was sich in Prompts nutzen lässt. Hier wäre interessant zu evaluieren, welche Prompt-Inhalte Reflexion direkt auslösen können. Zusätzlich scheint die Anzahl der Wörter relevant dafür zu sein, wie einfach sich Inhalte in einem Text transportieren lassen. Hier ist es ebenfalls interessant zu analysieren, ob sich diese Erkenntnisse in der Praxis in Prompts umsetzen lassen.



## 9 Anhang

### A Vorgehen und Grundlagen für die statistische Analyse in Teilstudie 2.1

Diese Teilstudie versucht, den möglichen Einfluss von Prompts auf die von den Crowdworkern gegebenen Antworten zu analysieren. Um diese Analyse zu ermöglichen, werden die Texte inhaltlich kodiert (siehe Abschnitte 3.4 und 5.1.2.4). In diesem Versuch ist demnach der Prompt die unabhängige Variable, deren Einfluss auf die Codes aus der Inhaltsanalyse, die den Inhalt des Textes repräsentieren, gemessen wird. Die reflexionsbezogenen Codes sind die abhängigen Variablen.

Diese reflexionsbezogenen Codes geben pro Dokument an, ob der entsprechende Code vorhanden ist. Dementsprechend ist dies eine dichotome kategorische<sup>27</sup> Variable. Die unterschiedlichen Prompts sind eine Nominal-Variable.

Da die Daten nicht kontinuierlich sind, können diese nicht normalverteilt sein. Diese beiden Faktoren (fehlende Normalverteilung und ordinale/nominale Variablen) schließen parametrische Verfahren, wie bspw. ANOVA aus (McHugh 2013). Aus diesem Grund wird im Folgenden zunächst der  $\chi^2$ -Test genutzt, um zu überprüfen, ob die demografischen Daten einen Einfluss auf die reflexionsbezogenen Codes haben. Da eine der Voraussetzungen für den  $\chi^2$ -Test ist, dass jeder Wert mindestens 5x vorhanden ist (Field 2009, 695–697) und dies nicht bei allen demografischen Daten gegeben ist, werden diese in Abschnitt 5.1.4 zu neuen Variablen gruppiert.

Zur Überprüfung, ob die Prompts oder die demografischen Daten einen Einfluss auf die reflexionsbezogenen Codes haben, wird eine binär logistische Regression durchgeführt. Durch die dichotome abhängige Variable eignet sich dieses Vorgehen (Pallant 2011, 168; Field 2009, 265). Im Folgenden werden die unterschiedlichen Annahmen für dieses Verfahren diskutiert.

Eine Annahme ist, dass die Daten der unabhängigen Variablen nicht linear sein dürfen (Field 2009, 273). Da ausschließlich kategorische Daten in die Analyse einfließen, ist diese Annahme erfüllt.

Die zweite Annahme ist, dass die Daten unabhängig sind. Dies bedeutet beispielsweise, dass keine Person zwei Mal an dem Versuchen teilgenommen hat (Field 2009, 273).

---

<sup>27</sup> Für eine Übersicht der Variablentypen siehe Field (2009, 8–10).

Aufgrund der Beschränkung im Studienaufbau ist diese Annahme ebenfalls erfüllt (siehe Abschnitt 5.1.2.1).

Die dritte Annahme ist, dass keine Multikollinearität besteht, sprich die unabhängigen Variablen untereinander nicht zu stark korrelieren (Field 2009, 273). Dies kann mithilfe von SPSS überprüft werden (Field 2009, 297–299). Zunächst werden dazu die VIF, *value inflated factors*, überprüft. Dieser Wert signalisiert, wie stark ein Prädiktor mit einem anderen Prädiktor zusammenhängt (Field 2009, 796). Je höher der Wert, desto eher liegt Multikollinearität vor. Bei den demografischen Werten in dieser Studie sind die Werte sehr niedrig (Tabelle 9-1), weswegen davon ausgegangen werden kann, dass keine Multikollinearität vorliegt. Die *Tolerance* errechnet sich aus dem VIF-Wert ( $1/\text{VIF}$ ) und ist problematisch, falls der Wert kleiner als 0,1 ist (Field 2009, 224).

Eine weitere Möglichkeit zur Überprüfung von Multikollinearität ist die Überprüfung der sogenannten Eigenvalues, welche die Verteilung der Varianz anzeigt (Field 2009, 243). Wenn der Eigenvalue über die Zeilen relativ ähnlich ist, bedeutet dies, dass sich das Regressionsmodell erst bei deutlichen Änderungen einer Variable ändert (Field 2009, 297–299). Eine Multikollinearität würde vorliegen, wenn bei den kleineren Werten bei Eigenvalue innerhalb einer Zeile zwei oder mehr Werte der einzelnen Variablen sehr hoch sind (Field 2009, 297–299). Es gibt hierbei keine Variablenkonstellation, in der die Variablen sehr stark miteinander korrelieren. In dieser Tabelle wurden bereits die in Abschnitt 5.1.4 gruppierten demografischen Daten überprüft.

**Tabelle 9-1: Value inflated factors zur Überprüfung der Multikollinearität in Teilstudie 2.1**

Model	Collinearity Statistics	
	Tolerance	VIF
Erfahrung > 1 Jahr	,860	1,163
Wochenarbeitsstunden	,977	1,024
Alter	,867	1,153
Geschlecht	,975	1,026
Akademiker (Bildung)	,978	1,023
Prompt	,975	1,026

**Tabelle 9-2: Eigenvalues zur Überprüfung der Multikollinearität in Teilstudie 2.1**

Dimension	Eigenvalue	Condition Index	Zeit	Muttersprache Englisch	Erfahrung > 1 Jahr	Wochenarbeitsstunden	Alter	Geschlecht	Akademiker (Bildung)	Prompts	Anzahl der Wörter
1	8,649	1,000	,00	,00	,00	,00	,00	,00	,00	,00	,00
2	,386	4,732	,09	,01	,00	,00	,02	,01	,01	,00	,55
3	,240	6,004	,51	,02	,01	,00	,07	,01	,03	,00	,23
4	,201	6,567	,15	,00	,12	,00	,39	,02	,00	,00	,02
5	,140	7,856	,05	,00	,01	,00	,01	,01	,10	,76	,00
6	,113	8,734	,01	,02	,17	,62	,11	,00	,02	,00	,02
7	,100	9,294	,07	,00	,05	,03	,03	,45	,29	,06	,05
8	,083	10,200	,00	,19	,35	,21	,13	,08	,23	,01	,01
9	,074	10,784	,12	,73	,00	,08	,00	,23	,07	,08	,06
10	,013	25,573	,00	,02	,30	,05	,23	,19	,25	,09	,07

In SPSS wird die Konfiguration für *forced entry* genutzt (anstatt *stepwise*), da in dieser Teilstudie Hypothesen getestet werden und die Daten nicht explorativ analysiert werden (vgl. im Folgenden Field 2009, 217–272). Bei forced entry werden alle Variablen direkt in ein Regressionsmodell eingebunden, wohingegen bei einem stepwise Modell sukzessive Variablen, die einen signifikanten Einfluss in dem Modell haben, hinzugefügt werden.

## **B Vorgehen und Grundlagen für die statistische Analyse in Teilstudie 2.2**

Diese Teilstudie versucht unter anderem, den möglichen Einfluss von Prompts auf die von den Crowdworkern gegebenen Antworten zu analysieren. Diese Teilstudie ist ähnlich angelegt, wie Teilstudie 2.1 (siehe Abschnitt 5.1). Dementsprechend ist die Beschreibung angelehnt an die vorherige Beschreibung in Anhang A.

Die Texte werden inhaltlich kodiert (siehe Abschnitte 3.4 und 5.1.2.4). Die unabhängige Variable ist der Prompt, dessen Einfluss auf die reflexionsbezogenen Codes aus der Inhaltsanalyse gemessen wird. Diese Codes stellen den Inhalt des Textes dar. Die reflexionsbezogenen Codes sind die abhängigen Variablen und geben pro Dokument an,

ob der entsprechende Code vorhanden ist. Dementsprechend ist dies eine dichotome kategorische<sup>28</sup> Variable. Die unterschiedlichen Prompts sind eine Nominal-Variable.

Viele statistische Verfahren basieren auf der Annahme, dass die Werte unabhängig voneinander sind, sprich u. a. von unterschiedlichen Probanden stammen (Field 2009, 273). Dies ist bspw. der Fall beim  $\chi^2$ -Test (Field 2009, 691), ANOVA (Field 2009, 360), und Regression (Field 2009, 273). Da diese Teilstudie auf der gleichen Plattform, wie die vorherige Teilstudie, durchgeführt wurde, gibt es mehrere Personen, die in beiden Studien teilgenommen haben. Für die Hypothesen 2 und 3, in denen nur Daten aus dieser Teilstudie analysiert werden ist dies nicht problematisch. Dies ist für Hypothese 1, in der die Antworten auf den Lösungs-Prompt zwischen Teilstudie 2.1 und 2.2 verglichen werden, problematisch. Es gibt insgesamt 130 Personen, die in beiden Studien teilgenommen haben. Allerdings gibt es nur 17 Personen, die in beiden Studien teilgenommen haben und in beiden Studien den Lösungs-Prompt bekommen haben. Da zur Überprüfung von Hypothese 1 nur diese 17 Personen problematisch sind, werden diese für die Analyse ausgeschlossen (siehe Abschnitt 5.2.5). Dies ist nur für Hypothese 1 relevant.

Für die Überprüfung der ersten Hypothese, die besagt, dass der Lösungs-Prompt bei einer vorhandenen Diskussion zu mehr Lösungsvorschlägen führt als bei einem neuen Thema, wird ein Vergleich zwischen zwei Gruppen benötigt: zwischen der Menge der Antworten auf den Prompt in einem Thema ohne Diskussion und der Menge der Antworten auf den Prompt in einem Thema mit Diskussion. Da die reflexionsbezogenen Codes dichotom sind, wird ein nicht-metrischer Test benötigt. Da die Personen überwiegend nur einmal an dem Versuch teilgenommen haben und der Datensatz (siehe oben) leichter dahingehend bereinigt werden kann, eignet sich zunächst ein  $\chi^2$ -Test (Field 2009, 691). Es wird ebenfalls eine binär logistische Regression durchgeführt, um zu prüfen, ob die Zugehörigkeit zu der einen (keine existierende Diskussion) oder anderen (existierende Diskussion) Gruppe eine Auswirkung auf die Codes hat.

Aus den bereits beschriebenen Gründen der fehlenden Normalverteilung und den ordinalen/nominalen Variablen werden parametrische Verfahren ausgeschlossen (siehe Anhang A). Zunächst werden daher  $\chi^2$ -Tests genutzt, um zu prüfen, ob sich die Verteilung der demografischen Daten von der Verteilung der reflexionsbezogenen Codes unterscheidet. Da der  $\chi^2$ -Test voraussetzt, dass jeder Wert mindestens 5x vorhanden ist (McHugh 2013), werden manche Variablen in Abschnitt 5.2.4 gruppiert.

---

<sup>28</sup> Für eine Übersicht der Variablentypen siehe Field (2009, 8–10).



Damit überprüft werden kann, inwieweit die Prompts oder die demografischen Daten einen Einfluss auf die reflexionsbezogenen Codes haben, wird eine logistische Regression durchgeführt. Dieses Verfahren eignet sich aufgrund der dichotomen abhängigen Variable (Pallant 2011, 168; Field 2009, 265). Die dafür notwendigen Annahmen werden im Folgenden diskutiert.

Die erste Annahme ist die Linearität, die besagt, dass Daten in unabhängigen Variablen nicht linear sein dürfen (Field 2009, 273). Die Annahme ist erfüllt, da ausschließlich kategoriale Daten in die Analyse einfließen.

Eine weitere Annahme ist, dass die Daten unabhängig sind, sprich von unterschiedlichen Personen stammen (Field 2009, 273). Der Datensatz wurde für die Überprüfung von Hypothese 1 bereinigt und für die weiteren Hypothesen ist diese Annahme bereits durch den Aufbau der Teilstudie erfüllt (siehe Abschnitt 5.2.2.1).

Die Abwesenheit von Multikollinearität, dem Zustand, dass die unabhängigen Variablen nicht miteinander korrelieren, ist die dritte Annahme (Field 2009, 273). Diese Annahme wurde mit SPSS überprüft (Field 2009, 297–299). Die Daten werden auf Multikollinearität überprüft. Das Vorgehen wurde zuvor für Teilstudie 2.1 in Anhang A beschrieben.

Tabelle 9-3 zeigt, dass die value inflated factors (VIF) für die unabhängigen Variablen alle nahe bei dem Wert eins liegen, was bedeutet, dass kein Verdacht auf Multikollinearität vorliegt. Die *Tolerance*-Werte liegen alle deutlich über 0,1 und bestätigen die erste Einschätzung.

**Tabelle 9-3: Value inflated factors zur Überprüfung der Multikollinearität in Teilstudie 2.2**

Model	Collinearity Statistics	
	Tolerance	VIF
Zeit	,798	1,253
Muttersprache Englisch	,895	1,118
Erfahrung > 1 Jahr	,921	1,086
Wochenarbeitsstunden	,938	1,066
Alter	,891	1,123
Geschlecht	,967	1,034
Akademiker (Bildung)	,977	1,024
Prompts	,962	1,040
Anzahl der Wörter	,837	1,195

Tabelle 9-4 zeigt, dass in den höheren Dimensionen keine Variablen vorhanden sind, die eine sehr hohen Korrelation untereinander aufweisen. Somit sollte in dieser Teilstudie die

Multikollinearität kein Problem darstellen. Für diese Kalkulation wurden die bereits gruppierte Variablen der demografischen Daten genutzt.

**Tabelle 9-4: Überprüfung der Multikollinearität in Teilstudie 2.1**

Dimension	Eigenvalue	Condition Index	Zeit	Muttersprache Englisch	Erfahrung > 1 Jahr	Wochenarbeitsstunden	Alter	Geschlecht	Akademiker (Bildung)	Prompts	Anzahl der Wörter
1	7,544	1,000	,00	,00	,00	,00	,00	,00	,00	,00	,00
2	0,538	3,744	,01	,14	,01	,04	,00	,39	,18	,00	,01
3	0,492	3,917	,01	,03	,02	,01	,00	,17	,62	,00	,01
4	0,408	4,302	,09	,40	,01	,02	,00	,23	,05	,00	,03
5	0,286	5,139	,09	,06	,19	,14	,02	,08	,02	,01	,27
6	0,252	5,472	,02	,12	,39	,48	,01	,02	,01	,00	,00
7	0,174	6,592	,01	,00	,25	,24	,16	,03	,02	,22	,17
8	0,160	6,865	,49	,15	,00	,00	,02	,00	,01	,22	,35
9	0,110	8,297	,20	,01	,08	,01	,70	,00	,00	,26	,13
10	0,037	14,214	,08	,09	,05	,04	,08	,08	,08	,29	,02

In SPSS wird die Konfiguration für *forced entry* genutzt (anstatt *stepwise*), da in dieser Teilstudie Hypothesen getestet werden und die Daten nicht explorativ analysiert werden (Field 2009, 272). Dies wurde bereits in Anhang A beschrieben.

### C Vorgehen und Grundlagen für die statistische Analyse in Teilstudie 2.3

In diesem Abschnitt wird beschrieben, welche statistischen Verfahren aufgrund der Beschaffenheit der Daten angewendet werden. Es wird zudem die Anwendbarkeit der Verfahren geprüft.

In dieser Teilstudie werden die Bewertungen von Lösungsvorschlägen aus zwei verschiedenen Gruppen, welche Ergebnisse aus Teilstudie 2.1 und 2.2 darstellen, verglichen. Die bewerteten Lösungsvorschläge bestehen aus mehreren Likert-Skalen mit jeweils fünf Kategorien. Die Likert-Skalen müssen als Ordinal-Variablen genutzt werden, da eine in numerische Werte umgewandelte Likert-Skala nicht automatisch eine Ratio/Intervall-Variable ist und die Abstände zwischen *strongly disagree* zu *disagree* zu *neither disagree nor agree* nicht äquidistant sind (Jamieson 2004). Da die Ergebnisse zwischen den zwei Teilstudien verglichen werden, wird der Unterschied von zwei Gruppen (unabhängige Variable) im Kontext der Bewertungen (abhängige Variablen)

verglichen. Die Probanden sind zwischen den beiden Gruppen unterschiedlich. Es gab zwar Personen, die an beiden Studien teilgenommen haben, aber es gibt keine Überlappung hinsichtlich der in dieser Teilstudie betrachteten erfahrungsbasierten Lösungsvorschläge (SUG\_EXP). Somit handelt es sich um zwei voneinander unabhängige Gruppen. Hierzu eignet sich ein Mann-Whitney-U-Test (Field 2009, 540 ff.). Die bei Field erwähnten Anforderungen sind demnach erfüllt (Field 2009, 540 ff.): das Vorhandensein einer kategorischen abhängigen Variablen mit mehreren Gruppen, eine dichotome unabhängige Variable und die Nicht-Erfüllung der Voraussetzungen für parametrische Verfahren. Es werden keine weiteren Annahmen erwähnt.

Tabelle 9-6 beinhaltet eine Analyse auf eine mögliche Multikollinearität der Daten. Sowohl die Werte für VIF als auch *Tolerance* zeigen keine Multikollinearität. Dies bedeutet, dass die unabhängigen Variablen in dieser Teilstudie nicht untereinander korrelieren (Field 2009, 273). Es wurden alle zur Verfügung stehenden demografischen Variablen untersucht: Dies beinhaltet sowohl die Daten, welche die Bewertenden angegeben haben und auch die Daten der Autoren der bewerteten Beiträge. Die Tabelle zeigt, dass aufgrund der Verteilung der Werte in den höchsten Dimensionen keine Multikollinearität vorliegt (Field 2009, 297–300). Eine Erklärung dieser Werte und Hintergründe befindet sich in Anhang A.

Zusätzlich hat ein Test ergeben, dass die Werte in allen Bewertungskategorien nicht normalverteilt sind (Tabelle 9-8), was ein weiteres Argument für einen nicht-parametrischen Test ist (Field 2009, 545–546).

Zur Analyse, inwieweit demografische Variablen einen Einfluss auf die Bewertung der Lösungsvorschläge nehmen, wird eine Regression durchgeführt. Da die Likert-Skalen ordinale Variablen sind, wird hier eine ordinale Regression benötigt. In das Regressionsmodell werden alle demografischen Daten der Autoren der Texte als auch der Bewertenden aufgenommen. Zusätzlich ist eine Variable enthalten, die markiert, ob die Texte aus Teilstudie 2.1 oder 2.2 stammen und so den Einfluss der Diskussionsbeiträge markiert. Weitere Variablen sind die zusammengefassten demografischen Variablen, wie die Altersgruppen oder die Erfahrung als Crowdworker.

In Tabelle 9-5 ist eine Übersicht der Eigenschaften der jeweiligen Regressionsmodelle. Die Spalten zu *Pseudo R-Square* beschreiben Indikatoren, welche Varianz die unabhängigen Variablen im ordinalen Regressionsmodell erklären (Denham 2017, 128–129). Dabei ist hervorzuheben, dass diese Werte bei einer ordinalen Regression mehr das Modell an sich erklären und sich nicht zum Vergleich mit anderen Modellen eignen (Denham 2017, 128–129). Cohen beschreibt Effektgrößen von ,10 als klein, ,30 als

mittelgroß und ,50 als groß (Cohen 1988, 77–81). Diese Werte beziehen sich allerdings auf eine lineare Regression und sind damit nur eingeschränkt mit den Effektgrößen der ordinalen Regression vergleichbar.

Die Spalten zu *Model Fitting* beschreiben, wie sehr sich das Modell mit den unabhängigen Variablen von dem Basismodell ohne die unabhängigen Variablen unterscheidet (Field 2009, 304). Ein signifikanter Wert bedeutet, dass das Hinzufügen der unabhängigen Variablen das Modell signifikant verbessert hat.

Die Spalten für *Goodness-of-Fit* beinhalten zwei unterschiedliche Metriken dafür, welche die Vorhersage des Modells bewerten. Der Test überprüft, ob sich die vorhergesagte Varianz der abhängigen Variable im Regressionsmodell von der beobachteten Varianz der Werte unterscheidet (Field 2009, 714–715). Dementsprechend werden hier nicht-signifikante Testergebnisse erwartet, welche dann signalisieren, dass die vorhergesagte Varianz des Regressionsmodells zu der beobachteten Varianz passt. Dies wird in beiden Tests (Pearson und Deviance) erwartet.

Eine Voraussetzung für dieses Verfahren ist, dass das Regressionsmodell für jede Ausprägung der ordinalen Variable (in diesem Fall jede Bewertungsstufe) funktioniert (Vgl. im Folgenden Denham 2017, 175). Diese Voraussetzung kann in SPSS mit dem Test of Parallel Lines durchgeführt werden. Wenn dieser Test ein nicht-signifikantes Ergebnis hat, dann sind die Voraussetzungen für eine ordinale Regression erfüllt. Falls dies nicht der Fall ist, dann ist nicht für jede Ausprägung der ordinalen Regression ein gültiges Modell vorhanden und die Berechnung muss mit einem weniger restriktiven Verfahren, wie bspw. einer nominalen Regression, durchgeführt werden.

Die Tabelle 9-5 zeigt, dass es lediglich für die beiden Kategorien der Glaubwürdigkeit eines Lösungsvorschlags (K7) und ob man den Lösungsvorschlag selbst nutzen würde (K12) ein gültiges ordinale Regressionsmodell gibt. Bei diesen Kategorien ist der Test of Parallel Lines erfüllt. Die letzte Kategorie, K12, wurde in einer 3er-Likert-Skala erhoben, was vermutlich erklärt, warum sich die Werte in den Goodness-of-Fit Tests von den anderen Kategorien unterscheiden.

Weiterhin ist erkennbar, dass für die Kriterien K1 und K3 das Regressionsmodell mit den unabhängigen Variablen zu keinem besseren Ergebnis kommt, was an den Ergebnissen in den Spalten für Model Fitting erkennbar ist (Tabelle 9-5).

Für die übrigen Kategorien (K2, K4, K5, K6, K8, K9, K10, K11) sind die Ergebnisse hinsichtlich Model Fitting und Goodness-of-Fit positiv, aber die Kategorien erfüllen die Voraussetzung des Test of Parallel Lines für die 5er-Likert-Skala nicht. Dies bedeutet,

dass es kein gültiges Regressionsmodell für jede Ausprägung der ordinalen abhängigen Variable existiert. Um gültige Regressionsmodelle für die ordinalen abhängigen Variablen zu entwickeln, wurde ebenfalls versucht, Regressionsmodelle auf Basis der reduzierten 3er-Likert-Skala zu berechnen (siehe Tabelle 9-7).

Aus der Tabelle ist ersichtlich, dass die Kategorien K2, K5, K7, K8, K9, K10, K11 sowohl die Voraussetzungen des Test of Parallel Lines bestehen als auch ein gültiges Modell aufweisen (siehe Tabelle 9-7).

Da für die Kategorie K7 bereits ein gültiges Modell auf Basis der 5er-Likert-Skala berechnet wurde, wird dieses verwendet. Für die übrigen Kategorien wird im Folgenden ein Regressionsmodell auf der reduzierten 3er-Likert-Skala genutzt. Im Vergleich zu einer multinominalen Regression, hat das ordinale Regressionsmodell den Vorteil, dass die Zusammenhänge für alle Ausprägungen der ordinalen Variable gelten.

Für die Kategorien K1 und K3 kam auch bei der reduzierten Likert-Skala kein gültiges Regressionsmodell zustande. Hier scheint es keine Variablen zu geben, welche die Varianz der Bewertungskategorien erklären. Hier schlägt bereits der Test für Model Fitting fehlt.

Für die Kategorien K4 und K6 gibt es keine gültigen Regressionsmodelle für eine ordinale Regression gemäß des Test of Parallel Lines. Dementsprechend gibt es keine Modelle, in denen für jede Ausprägung der unabhängigen Variable ein gültiges Modell existiert. Hier sind weitere Analysen notwendig, welche den möglichen Einfluss von demografischen Variablen analysieren.

Die Regressionsmodelle sind in Anhang 0 dargestellt. Dort wird geprüft, welche Faktoren in jeder Ausprägung signifikant zum Modell beitragen. Von diesen signifikant zum Modell beitragenden Faktoren werden die Wahrscheinlichkeiten für eine der Bewertungskategorien (*strongly disagree* bis *strongly agree*) berechnet.

Alle Modelle haben in SPSS eine Warnung, dass ein hoher Wert (zwischen 65 % und 67 %) an Ausprägungskombinationen der unabhängigen Variablen fehlt (genannt „zero frequencies“) (Field 2009, 307). Bei der Regressionsanalyse wird darauf geachtet, dass für jede mögliche Kombination aller Ausprägungen aller unabhängigen Variablen ein Eintrag in der Datenmenge vorhanden ist. Da hier eine hohe Anzahl Daten fehlt, sind die Daten nur eingeschränkt interpretierbar. In der Literatur finden sich allerdings auch Studien, die mit deutlich höheren Werten gearbeitet haben (bspw. 75 % bei Polyzos und Minetos 2011; 80-85 % bei Spais und Vasileiou 2006; und 80 % bei Ellena u. a. 2016). Letztlich lässt sich dieses Problem auf zwei Arten beheben: 1) mehr Daten erheben und

darauf achten auch spezifische seltene Kombinationen an Ausprägungen zu bekommen, oder 2) Regressionsmodelle mit weniger Variablen erstellen. Bei Ersterem ist das Problem, dass bspw. nicht alle Altersgruppe oder alle Bildungsgruppen gleich leicht zu akquirieren sind. Bei Letzterem ist das Problem, dass möglicherweise wichtige Variablen fehlen. Bei Variablen, die sehr viele Ausprägungen haben (bspw. Alter und Wochenarbeitsstunden), ist auch in Modellen mit sehr wenigen Variablen bereits ein höherer Wert an fehlenden Ausprägungskombinationen vorhanden (34 % in einem Modell für K7 bestehend aus der Studie, den Wochenarbeitsstunden, Alter und Geschlecht des Autors). Aus diesem Grund wurden Modelle mit allen Variablen berechnet, die allerdings nur vorsichtig interpretiert werden dürfen.

**Tabelle 9-5: Übersicht der Eigenschaften der ordinalen Regressionsmodelle (5er-Likert-Skala), die aus allen demografischen Daten (sowohl der Autoren der Beiträge als auch der Bewertenden) sowie einer Variable für die Diskussionsbeiträge (Unterscheidung Teilstudien 2.1 und 2.2) bestehen. Es wurden Regressionsmodelle für alle Kategorien und für alle Texte berechnet, die bei einem Lösungs-Prompt entstanden sind. Valide Regressionsmodelle sind mit \* markiert.**

Kategorie	Pseudo R-Square			Model Fitting (df=31)		Goodness-of-Fit (df=1225)				Test of Parallel Lines (df=93)	
	Cox and Snell	Nagel- kerke	McFadden	Chi- Square	Sig.	Pearson		Deviance		Chi- Square	Sig.
						Chi- Square	Sig.	Chi- Square	Sig.		
Novel (K1)	,126	,133	,046	44,036	,061	1276,112	,151	896,182	1	111,154	,097
Unusual (K2)	,132	,142	,053	46,239	,039	1205,506	,649	806,978	1	141,538	,001
Elements I did not think of (K3)	,121	,131	,049	42,074	,089	1310,525	,044	790,045	1	99,764	,297
Fits the issue (K4)	,226	,242	,095	83,366	<,001	1256,329	,261	771,069	1	201,727	<,001
Likely solves the issue (K5)	,190	,200	,071	68,832	<,001	1206,881	,638	882,865	1	209,661	<,001
Helpful (K6)	,217	,230	,086	79,769	<,001	1272,202	,170	825,157	1	148,244	<,001
Believable (K7)*	,204	,224	,094	74,277	<,001	1161,199	,903	695,924	1	77,972	,868
Clearly stated (K8)	,245	,262	,103	91,818	<,001	1146,750	,945	777,770	1	135,611	,003
Put into practice (K9)	,215	,229	,088	78,712	<,001	1177,622	,831	789,734	1	137,705	,002
Impact (K10)	,173	,182	,064	61,945	,001	1255,050	,269	878,649	1	119,618	,033
Recommend (K11)	,234	,246	,089	86,741	<,001	1304,531	,056	861,837	1	125,275	,014
I would use it (K12)*	,191	,220	,105	69,121	<,001	607,181 (df=597)	,377	573,678 (df=578)	,747	37,505 (df=31)	,195

**Tabelle 9-6: Überprüfung der Multikollinearität in Teilstudie 2.3. In der Tabelle werden alle demografischen Daten sowohl der Autoren (erhoben in Teilstudie 2.2) als auch der Bewertenden (erhoben in Teilstudie 2.3) überprüft.**

		Demografie der Bewertenden							Demografie der Autoren							
Dimension	Eigenvalue	Condition Index	Wochenarbeits- stunden	Muttersprache Englisch	Erfahrung > 1 Jahr	Alter	Geschlecht	Akademiker (Bildung)	Anzahl Wörter	Zeit	Muttersprache Englisch	Erfahrung > 1 Jahr	Wochenarbeits- stunden	Alter	Geschlecht	Akademiker (Bildung)
1	11,880	1,000	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00	,00
2	,603	4,437	,00	,01	,00	,00	,68	,01	,00	,00	,00	,00	,00	,00	,00	,01
3	,540	4,691	,00	,01	,00	,00	,00	,73	,00	,00	,01	,00	,00	,00	,00	,04
4	,464	5,059	,00	,00	,00	,00	,03	,11	,00	,00	,00	,03	,00	,00	,01	,48
5	,337	5,933	,00	,00	,00	,00	,00	,00	,00	,17	,36	,02	,01	,00	,00	,01
6	,257	6,793	,00	,03	,01	,01	,03	,01	,00	,08	,00	,09	,06	,14	,00	,03
7	,210	7,527	,01	,00	,00	,02	,00	,02	,00	,17	,35	,31	,01	,00	,00	,01
8	,173	8,294	,20	,08	,16	,10	,00	,06	,00	,00	,00	,01	,00	,02	,01	,03
9	,137	9,308	,37	,16	,00	,04	,01	,00	,03	,01	,00	,16	,00	,02	,02	,09
10	,114	10,218	,03	,13	,00	,16	,00	,01	,12	,05	,00	,01	,02	,09	,12	,08
11	,092	11,362	,00	,32	,07	,41	,07	,01	,00	,00	,06	,07	,20	,02	,00	,04
12	,073	12,769	,16	,03	,33	,18	,00	,00	,02	,02	,04	,00	,41	,06	,02	,03
13	,067	13,267	,00	,01	,03	,03	,01	,00	,09	,31	,10	,28	,01	,61	,16	,01
14	,045	16,314	,11	,11	,24	,01	,07	,00	,47	,19	,00	,02	,01	,00	,39	,12
15	,007	40,434	,11	,12	,14	,04	,10	,04	,27	,01	,06	,01	,26	,03	,27	,00
Tolerance			,861	,796	,779	,871	806	,925	,662	,659	,823	,749	,793	,618	,579	,748
VIF			1,162	1,257	1,284	1,148	1,240	1,082	1,511	1,517	1,215	1,335	1,261	1,618	1,726	1,337



Tabelle 9-7: Übersicht der Eigenschaften der ordinalen Regressionsmodelle (3er-Likert-Skala), die aus allen demografischen Daten (sowohl der Autoren der Beiträge als auch der Bewertenden) sowie einer Variable für die Diskussionsbeiträge (Unterscheidung Teilstudien 2.1 und 2.2) bestehen. Es wurden Regressionsmodelle für alle Kategorien und für alle Texte berechnet, die bei einem Lösungs-Prompt entstanden sind. Valide Regressionsmodelle sind mit \* markiert.

Kategorie	Pseudo R-Square			Model Fitting (df=31)		Goodness-of-Fit (df=597)				Test of Parallel Lines (df=31)	
	Cox and Snell	Nagel- kerke	McFadden	Chi- Square	Sig.	Pearson Chi- Square	Sig.	Deviance Chi- Square	Sig.	Chi-Square	Sig.
<b>Novel (K1)</b>	,127	,145	,064	44,462	,056	636,181	,129	641,618	,122	47,553	,029
<b>Unusual (K2)*</b>	,134	,164	,086	46,736	,035	653,330	,055	491,494	,999	17,619	,974
<b>Elements I did not think of (K3)</b>	,105	,129	,067	36,060	,244	626,756	,193	490,608	,999	15,449	,991
<b>Fits the issue (K4)</b>	,244	,306	,175	91,393	<,001	589,703	,576	414,714	1	75,388	<,001
<b>Likely solves the issue (K5)*</b>	,205	,231	,105	74,857	<,001	619,660	,252	616,432	,282	26,493	,697
<b>Helpful (K6)</b>	,238	,286	,152	88,721	<,001	628,750	,178	479,652	1	74,286	<,001
<b>Believable (K7)*</b>	,237	,321	,202	88,095	<,001	614,936	,297	333,991	1	27,750	,634
<b>Clearly stated (K8)*</b>	,264	,328	,187	100,082	<,001	552,677	,903	416,337	1	42,286	,085
<b>Put into practice (K9) *</b>	,221	,278	,158	81,234	<,001	628,873	,177	417,453	1	44,323	,057
<b>Impact (K10)*</b>	,181	,207	,096	65,141	<,001	624,736	,209	591,429	,559	32,503	,393
<b>Recommend (K11)*</b>	,239	,282	,145	89,703	<,001	615,009	,296	511,837	,995	9,719	1

**Tabelle 9-8: Test auf Normalverteilung der Bewertungskategorien.**

Kategorie	Teil- studie	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Novel (K1)	2.1	,202	369	,000	,907	369	,000
	2.2	,219	384	,000	,900	384	,000
Unusual (K2)	2.1	,289	369	,000	,845	369	,000
	2.2	,274	384	,000	,871	384	,000
Elements I didn't think of (K3)	2.1	,292	369	,000	,857	369	,000
	2.2	,313	384	,000	,847	384	,000
Fits the issue (K4)	2.1	,317	369	,000	,829	369	,000
	2.2	,290	384	,000	,854	384	,000
Solves the issue (K5)	2.1	,203	369	,000	,907	369	,000
	2.2	,185	384	,000	,908	384	,000
Helpful (K6)	2.1	,288	369	,000	,863	369	,000
	2.2	,259	384	,000	,879	384	,000
Believable (K7)	2.1	,322	369	,000	,793	369	,000
	2.2	,314	384	,000	,819	384	,000
Clearly stated (K8)	2.1	,303	369	,000	,830	369	,000
	2.2	,275	384	,000	,851	384	,000
Put into practice (K9)	2.1	,302	369	,000	,827	369	,000
	2.2	,276	384	,000	,858	384	,000
Impact (K10)	2.1	,255	369	,000	,887	369	,000
	2.2	,223	384	,000	,898	384	,000
Recommend (K11)	2.1	,288	369	,000	,864	369	,000
	2.2	,244	384	,000	,883	384	,000

## D Statistiken für Teilstudie 2.3 – Demografische Daten

Tabelle 9-9: Zusammenhang zwischen Bewertungskriterien und der Variable Academics.

Rating	Var	$\chi^2$ (df=4)	Strongly disagree	disagree	Neither nor	agree	Strongly agree
<b>Recommend (K11)</b>	No Acad.	$\chi^2=29,174$ ,	14,2 %	11,6 %	14,5 %	37,9 %	21,8 %
	Acad.	$p<,001$	6,6 %	20,7 %	17,3 %	41,7 %	13,6 %
<b>Impact (K10)</b>	No Acad.	$\chi^2=11,245$ ,	11,8 %	19,9 %	19,9 %	35,2 %	13,2 %
	Acad.	$p=,024$	8,1 %	18,6 %	24,9 %	40,4 %	7,9 %
<b>Put into practice (K9)</b>	No Acad.	$\chi^2=22,072$ ,	7,8 %	9,9 %	12,9 %	39,8 %	29,6 %
	Acad.	$p<,001$	2,6 %	9,2 %	17,6 %	49,6 %	21 %
<b>Fits the issue (K4)</b>	No Acad.	$\chi^2=16,328$ ,	3,8 %	12,1 %	10,8 %	46,8 %	26,6 %
	Acad.	$p=,003$	3,4 %	9,4 %	18,6 %	50,7 %	17,8 %
<b>Suggestion is Believable (K7)</b>	No Acad.	$\chi^2=15,404$ ,	3,5 %	7,5 %	12,9 %	46,8 %	29,3 %
	Acad.	$p=,004$	1 %	5,8 %	10,8 %	59,6 %	22,8 %
<b>Elements I Didn't think of ()</b>	No Acad.	$\chi^2=31,830$ ,	26,3 %	40,9 %	15,1 %	13,7 %	4 %
	Acad.	$p<,001$	11,5 %	54,6 %	13,6 %	17,3 %	2,9 %
<b>Clearly stated (K8)</b>	No Acad.	$\chi^2=18,445$ ,	5,9 %	12,4 %	13,4 %	38,4 %	29,8 %
	Acad.	$p=,001$	2,9 %	7,3 %	15,2 %	50,9 %	23,6 %
<b>Helpful (K6)</b>	No Acad.	$\chi^2=15,285$ ,	7,3 %	12,9 %	15,3 %	40,9 %	23,7 %
	Acad.	$p=,004$	3,1 %	14,7 %	21 %	44,6 %	16,5 %
<b>Novel (K1)</b>	No Acad.	$\chi^2=21,461$ ,	12,6 %	30,9 %	19,1 %	27,4 %	9,9 %
	Acad.	$p<,001$	10 %	34,4 %	27,6 %	24,9 %	3,1 %
<b>Unusual (K2)</b>	No Acad.	$\chi^2=18,342$ ,	28,2 %	44,1 %	11,6 %	12,6 %	3,5 %
	Acad.	$p=,001$	18,9 %	44,4 %	20,2 %	14,7 %	1,8 %

**Tabelle 9-10: Zusammenhang je Bewertung und der Anzahl der Wörter (3er-Likert-Skala)**

<b>Rating</b>	<b>Wörter</b>	<b><math>\chi^2</math> (df=6)</b>	<b>negative</b>	<b>Neither nor</b>	<b>positive</b>
<b>Recommend (K11)</b>	0-40	$\chi^2=33,240$ , p<,001	60,0 %	8,9 %	31,1 %
	41-80		27,7 %	17,9 %	54,4 %
	81-120		22,2 %	13,9 %	63,9 %
	121+		21,6 %	17,0 %	61,4 %
<b>Impact (K10)</b>	0-40	$\chi^2=25,090$ , p<,001	60,0 %	17,8 %	22,2 %
	41-80		28,7 %	23,1 %	48,2 %
	81-120		27,4 %	23,9 %	48,7 %
	121+		24,6 %	20,5 %	55,0 %
<b>Put into practice (K9)</b>	0-40	$\chi^2=26,813$ , p<,001	14,4 %	9,6 %	3,4 %
	41-80		16,0 %	15,6 %	68,4 %
	81-120		12,2 %	14,3 %	73,5 %
	121+		10,5 %	13,5 %	76,0 %
<b>Likely solves the issue (K5)</b>	0-40	$\chi^2=29,871$ , p<,001	55,6 %	28,9 %	15,6 %
	41-80		35,2 %	31,6 %	33,2 %
	81-120		25,2 %	30,0 %	44,8 %
	121+		25,7 %	26,3 %	48,0 %
<b>Fits the issue (K4)</b>	0-40	$\chi^2=33,512$ , p<,001	40,0 %	17,8 %	42,2 %
	41-80		14,0 %	17,3 %	68,7 %
	81-120		12,2 %	13,9 %	73,9 %
	121+		11,1 %	10,5 %	78,4 %
<b>Suggestion is Believable (K7)</b>	0-40	$\chi^2=46,475$ , p<,001	33,3 %	20,0 %	46,7 %
	41-80		6,5 %	14,3 %	79,2 %
	81-120		8,3 %	9,6 %	82,2 %
	121+		7,6 %	8,2 %	84,2 %
<b>Clearly stated (K8)</b>	0-40	$\chi^2=26,311$ , p<,001	31,1 %	24,4 %	44,4 %
	41-80		13,7 %	16,9 %	69,4 %
	81-120		14,8 %	12,6 %	72,6 %
	121+		9,9 %	9,4 %	80,7 %
<b>Helpful (K6)</b>	0-40	$\chi^2=39,521$ , p<,001	48,9 %	28,9 %	22,2 %
	41-80		18,6 %	18,2 %	63,2 %
	81-120		16,5 %	17,4 %	66,1 %
	121+		15,2 %	16,4 %	68,4 %
<b>Novel (K1)</b>	0-40	$\chi^2=16,653$ , p=,016	57,8 %	8,9 %	33,3 %
	41-80		46,9 %	26,4 %	26,7 %
	81-120		39,6 %	21,7 %	38,7 %
	121+		40,9 %	24,0 %	35,1 %
<b>Rating</b>	<b>Wörter</b>	<b><math>\chi^2</math> (df=6)</b>	<b>no</b>	<b>maybe</b>	<b>yes</b>
<b>Would you use it (K12)</b>	0-40	$\chi^2=28,759$ , p<,001	55,6 %	22,2 %	22,2 %
	41-80		24,1 %	26,4 %	49,5 %
	81-120		25,2 %	21,3 %	53,5 %
	121+		21,1 %	21,1 %	57,9 %

**Tabelle 9-11: Zusammenhang zwischen der Variable für die Wochenstunden der Autoren  
und den Bewertungskriterien**

<b>Rating</b>	<b>Wochenstunden</b>	<b><math>\chi^2</math> (df=8)</b>	<b>negative</b>	<b>Neither nor</b>	<b>positive</b>
<b>Fits the issue (K4)</b>	< 1 Stunde	$\chi^2=19,971$ , p=,010	22,2 %	25,4 %	52,4 %
	1-5 Stunden		11,8 %	12,9 %	75,3 %
	5-15 Stunden		15,9 %	14,1 %	70 %
	15-30 Stunden		15,1 %	9,4 %	75,5 %
	> 30 Stunden		15,9 %	25 %	59,1 %
<b>Likely solves the issue (K5)</b>	< 1 Stunde	$\chi^2=19,552$ , p=,012	49,2 %	22,2 %	28,6 %
	1-5 Stunden		29,8 %	30,3 %	39,9 %
	5-15 Stunden		25,9 %	29,1 %	45 %
	15-30 Stunden		35,8 %	30,2 %	34 %
	> 30 Stunden		38,6 %	38,6 %	22,7 %
<b>Helpful (K6)</b>	< 1 Stunde	$\chi^2=23,836$ , p=,002	30,2 %	12,7 %	57,1 %
	1-5 Stunden		15,8 %	16,9 %	67,3 %
	5-15 Stunden		16,4 %	20,5 %	63,2 %
	15-30 Stunden		30,2 %	15,1 %	54,7 %
	> 30 Stunden		29,5 %	29,5 %	40,9 %
<b>Clearly stated (K8)</b>	< 1 Stunde	$\chi^2=38,626$ , p<,001	31,7 %	15,9 %	52,4 %
	1-5 Stunden		11,3 %	13,4 %	75,3 %
	5-15 Stunden		11,8 %	12,7 %	75,5 %
	15-30 Stunden		9,4 %	20,8 %	69,8 %
	> 30 Stunden		31,8 %	20,5 %	47,7 %
<b>Put into practice (K9)</b>	< 1 Stunde	$\chi^2=24,442$ , p=,002	28,6 %	17,5 %	54 %
	1-5 Stunden		11,5 %	15,8 %	72,7 %
	5-15 Stunden		14,1 %	11,8 %	74,1 %
	15-30 Stunden		15,1 %	15,1 %	69,8 %
	> 30 Stunden		25 %	25 %	50 %
<b>Impact (K10)</b>	< 1 Stunde	$\chi^2=26,224$ , p=,001	42,9 %	22,2 %	34,9 %
	1-5 Stunden		23,6 %	24,9 %	51,5 %
	5-15 Stunden		28,6 %	19,5 %	51,8 %
	15-30 Stunden		35,8 %	22,6 %	41,5 %
	> 30 Stunden		52,3 %	15,9 %	31,8 %
<b>Recommend (K11)</b>	< 1 Stunde	$\chi^2=30,401$ , p<,001	36,5 %	27 %	36,5 %
	1-5 Stunden		22 %	15 %	63 %
	5-15 Stunden		25,5 %	13,6 %	60,9 %
	15-30 Stunden		35,8 %	18,9 %	45,3 %
	> 30 Stunden		45,5 %	15,9 %	38,6 %

## E Ordinale und Multinominale Regressionsmodelle für Teilstudie 2.3

Tabelle 9-12: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie ungewöhnlich der Lösungsvorschlag ist (K2). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-7 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet.

		Estimate	Std. Error	Wald	df	Sig.	95% Conf. Interval	
							Lower Bound	Upper Bound
Thr	Neutral Bewertung	3,299	1,261	6,846	1	,009	,828	5,771
	Positive Bewertung	4,314	1,272	11,499	1	,001	1,821	6,808
Location	Teilstudie 2.1	-,258	,321	,648	1	,421	-,886	,370
	Teilstudie 2.2	0 <sup>a</sup>	.	.	0	.	.	.
	Wörter: 0-40	1,520	1,105	1,893	1	,169	-,645	3,685
	Wörter: 41-80	,514	,382	1,817	1	,178	-,233	1,262
	Wörter: 81-120	,199	,442	,203	1	,652	-,667	1,066
	Wörter: 121+	0 <sup>a</sup>	.	.	0	.	.	.
	Zeit: 0-240s	-,470	,688	,467	1	,494	-1,819	,878
	Zeit: 241-480s	,135	,408	,109	1	,741	-,665	,934
	Zeit: 481-720s	-,944	,442	4,564	1	,033	-1,810	-,078
	Zeit: >720s	0 <sup>a</sup>	.	.	0	.	.	.
	B: Weiblich	,053	,287	,034	1	,854	-,509	,615
	B: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	B: Alter 17-24	-1,430	,850	2,827	1	,093	-3,096	,237
	B: Alter 25-34	,521	,395	1,741	1	,187	-,253	1,295
	B: Alter 35-44	,558	,416	1,792	1	,181	-,259	1,374
	B: Alter 45-54	,557	,411	1,837	1	,175	-,248	1,361
	B: Alter 55+	0 <sup>a</sup>	.	.	0	.	.	.
	B: Kein Akademiker	,130	,275	,224	1	,636	-,409	,669
	B: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	B: Keine Muttersprache Englisch	-,048	,387	,015	1	,902	-,807	,712
	B: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
	B: Erfahrung < 1 Jahr	,183	,524	,121	1	,728	-,844	1,209
	B: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
	B: < 1 Std. / Woche	,509	,451	1,272	1	,259	-,375	1,393
	B: 1-5 Std. / Woche	-,157	,366	,184	1	,668	-,875	,561
	B: 5-15 Std. / Woche	,270	,421	,411	1	,522	-,555	1,095
	B: > 15 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
	A: Weiblich	,103	,368	,078	1	,780	-,619	,824
	A: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	A: Alter 17-24	-3,059	1,204	6,460	1	,011	-5,419	-,700
	A: Alter 25-34	,224	,693	,105	1	,746	-1,134	1,582
	A: Alter 35-44	-1,057	,559	3,576	1	,059	-2,153	,039
	A: Alter 45-54	-,761	,477	2,548	1	,110	-1,695	,173
	A: Alter 55-64	-,877	,486	3,256	1	,071	-1,830	,076
	A: Alter 65+	0 <sup>a</sup>	.	.	0	.	.	.
	A: Kein Akademiker	-,280	,324	,749	1	,387	-,915	,354
	A: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.

	Estimate	Std. Error	Wald	df	Sig.	95% Conf. Interval	
						Lower Bound	Upper Bound
A: Keine Muttersprache Englisch	,768	,335	5,250	1	,022	,111	1,425
A: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
A: Erfahrung < 1 Jahr	,941	,413	5,202	1	,023	,132	1,749
A: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
A: < 1 Std. / Woche	2,500	1,267	3,891	1	,049	,016	4,984
A: 1-5 Std. / Woche	2,650	,970	7,459	1	,006	,748	4,552
A: 5-15 Std. / Woche	2,154	,958	5,056	1	,025	,276	4,032
A: 15-30 Std. / Woche	2,552	,994	6,598	1	,010	,605	4,499
A: >30 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
Link function: Logit.							
a. This parameter is set to zero because it is redundant.							

**Tabelle 9-13: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie sehr der Lösungsvorschlag das Problem lösen kann (K5). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 79 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet.**

		Estimate	Std. Error	Wald	d f	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Thr	Neutral Bewertung	-3,069	1,230	6,227	1	,013	-5,480	-,659
	Positive Bewertung	-1,592	1,221	1,700	1	,192	-3,986	,801
Location	Teilstudie 2.1	,781	,281	7,719	1	,005	,230	1,332
	Teilstudie 2.2	0 <sup>a</sup>	.	.	0	.	.	.
	Wörter: 0-40	-5,063	1,672	9,175	1	,002	-8,339	-1,787
	Wörter: 41-80	-,569	,344	2,739	1	,098	-1,243	,105
	Wörter: 81-120	,014	,389	,001	1	,970	-,748	,777
	Wörter: 121+	0 <sup>a</sup>	.	.	0	.	.	.
	Zeit: 0-240s	,339	,639	,281	1	,596	-,914	1,591
	Zeit: 241-480s	-,289	,362	,638	1	,424	-,999	,420
	Zeit: 481-720s	,162	,381	,181	1	,670	-,584	,908
	Zeit: >720s	0 <sup>a</sup>	.	.	0	.	.	.
	B: Weiblich	-,128	,257	,248	1	,619	-,631	,375
	B: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	B: Alter 17-24	,709	,514	1,901	1	,168	-,299	1,717
	B: Alter 25-34	,430	,337	1,632	1	,201	-,230	1,090
	B: Alter 35-44	-,273	,355	,591	1	,442	-,968	,423
	B: Alter 45-54	-,183	,353	,270	1	,603	-,875	,508
	B: Alter 55+	0 <sup>a</sup>	.	.	0	.	.	.
	B: Kein Akademiker	-,353	,251	1,980	1	,159	-,844	,139
	B: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	B: Keine Muttersprache Englisch	-,307	,362	,721	1	,396	-1,017	,402
	B: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
	B: Erfahrung < 1 Jahr	,012	,440	,001	1	,979	-,851	,875
	B: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
	B: < 1 Std. / Woche	-,299	,413	,523	1	,470	-1,109	,511
	B: 1-5 Std. / Woche	-,353	,314	1,262	1	,261	-,969	,263
	B: 5-15 Std. / Woche	-,588	,368	2,553	1	,110	-1,308	,133
	B: > 15 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
	A: Weiblich	-,569	,331	2,955	1	,086	-1,218	,080
	A: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	A: Alter 17-24	,448	,818	,300	1	,584	-1,156	2,051
	A: Alter 25-34	1,732	,658	6,927	1	,008	,442	3,022
	A: Alter 35-44	,435	,489	,793	1	,373	-,523	1,393
	A: Alter 45-54	-,241	,436	,305	1	,581	-1,095	,613
	A: Alter 55-64	-,070	,438	,025	1	,873	-,929	,789
	A: Alter 65+	0 <sup>a</sup>	.	.	0	.	.	.
	A: Kein Akademiker	-,599	,290	4,264	1	,039	-1,167	-,030
	A: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	A: Keine Muttersprache Englisch	-,519	,305	2,892	1	,089	-1,116	,079



	Estimate	Std. Error	Wald	d f	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
A: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
A: Erfahrung < 1 Jahr	-1,370	,393	12,173	1	,000	-2,139	-,600
A: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
A: < 1 Std. / Woche	-,801	1,179	,461	1	,497	-3,112	1,511
A: 1-5 Std. / Woche	-,585	,970	,364	1	,546	-2,486	1,316
A: 5-15 Std. / Woche	-,939	,944	,988	1	,320	-2,790	,912
A: 15-30 Std. / Woche	-1,375	,973	1,996	1	,158	-3,283	,533
A: >30 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
Link function: Logit.							
a. This parameter is set to zero because it is redundant.							

**Tabelle 9-14: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie glaubwürdig der Lösungsvorschlag ist (K7). Der Test für die Validität des Regressionsmodells auf Basis der 5er-Likert-Skala ist in Tabelle 9-5 abgebildet. Der Referenzwert für den Vergleich ist die Bewertung *strongly disagree*. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet.**

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	Disagree	-7,164	1,183	36,660	1	,000	-9,484	-4,845
	Neither nor	-5,279	1,088	23,531	1	,000	-7,412	-3,146
	Agree	-4,223	1,069	15,614	1	,000	-6,318	-2,128
	Strongly Agree	-1,392	1,048	1,765	1	,184	-3,446	,662
Location	Teilstudie 2.1	,701	,285	6,029	1	,014	,141	1,260
	Teilstudie 2.2	0 <sup>a</sup>	.	.	0	.	.	.
	Wörter: 0-40	-3,129	,962	10,570	1	,001	-5,015	-1,243
	Wörter: 41-80	-,261	,342	,584	1	,445	-,931	,408
	Wörter: 81-120	-,118	,390	,091	1	,762	-,881	,646
	Wörter: 121+	0 <sup>a</sup>	.	.	0	.	.	.
	Zeit: 0-240s	-,226	,606	,139	1	,710	-1,414	,962
	Zeit: 241-480s	-,675	,370	3,327	1	,068	-1,400	,050
	Zeit: 481-720s	-,538	,385	1,955	1	,162	-1,292	,216
	Zeit: >720s	0 <sup>a</sup>	.	.	0	.	.	.
	B: Weiblich	,205	,252	,663	1	,416	-,289	,699
	B: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	B: Alter 17-24	,277	,510	,295	1	,587	-,723	1,278
	B: Alter 25-34	-,556	,339	2,691	1	,101	-1,221	,108
	B: Alter 35-44	-1,080	,364	8,804	1	,003	-1,794	-,367
	B: Alter 45-54	-,687	,360	3,641	1	,056	-1,392	,019
	B: Alter 55+	0 <sup>a</sup>	.	.	0	.	.	.
	B: Kein Akademiker	-,398	,249	2,563	1	,109	-,885	,089
	B: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	B: Keine Muttersprache Englisch	-,010	,355	,001	1	,978	-,706	,686
	B: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
	B: Erfahrung < 1 Jahr	-,120	,428	,079	1	,779	-,959	,719
	B: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
	B: < 1 Std. / Woche	,230	,408	,318	1	,573	-,569	1,029
	B: 1-5 Std. / Woche	,261	,308	,721	1	,396	-,342	,864
	B: 5-15 Std. / Woche	-,082	,365	,051	1	,822	-,798	,634
	B: > 15 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
	A: Weiblich	-,161	,320	,254	1	,614	-,787	,465
	A: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	A: Alter 17-24	,913	,836	1,192	1	,275	-,726	2,552
	A: Alter 25-34	1,314	,627	4,401	1	,036	,086	2,542
	A: Alter 35-44	-,036	,492	,005	1	,941	-1,000	,927

	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
A: Alter 45-54	-,039	,440	,008	1	,929	-,902	,824
A: Alter 55-64	-,028	,443	,004	1	,949	-,896	,840
A: Alter 65+	0 <sup>a</sup>	.	.	0	.	.	.
A: Kein Akademiker	-,949	,292	10,579	1	,001	-1,522	-,377
A: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
A: Keine Muttersprache Englisch	-,463	,293	2,498	1	,114	-1,038	,111
A: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
A: Erfahrung < 1 Jahr	-1,094	,371	8,678	1	,003	-1,822	-,366
A: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
A: < 1 Std. / Woche	-,879	1,056	,693	1	,405	-2,948	1,191
A: 1-5 Std. / Woche	-,981	,773	1,610	1	,204	-2,495	,534
A: 5-15 Std. / Woche	-1,424	,769	3,430	1	,064	-2,931	,083
A: 15-30 Std. / Woche	-1,066	,814	1,715	1	,190	-2,662	,529
A: >30 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
Link function: Logit.							
a. This parameter is set to zero because it is redundant.							

**Tabelle 9-15: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie klar der Lösungsvorschlag formuliert ist (K8). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-7 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet.**

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Thr	Neutral Bewertung	-2,945	1,346	4,784	1	,029	-5,584	-,306
	Positive Bewertung	-1,557	1,332	1,367	1	,242	-4,169	1,054
Location	Teilstudie 2.1	1,190	,369	10,393	1	,001	,467	1,913
	Teilstudie 2.2	0 <sup>a</sup>	.	.	0	.	.	.
	Wörter: 0-40	-4,138	1,308	10,010	1	,002	-6,702	-1,575
	Wörter: 41-80	,060	,446	,018	1	,894	-,814	,933
	Wörter: 81-120	-,604	,505	1,432	1	,231	-1,594	,386
	Wörter: 121+	0 <sup>a</sup>	.	.	0	.	.	.
	Zeit: 0-240s	-,427	,741	,333	1	,564	-1,879	1,024
	Zeit: 241-480s	,669	,479	1,949	1	,163	-,270	1,608
	Zeit: 481-720s	,289	,494	,342	1	,559	-,680	1,258
	Zeit: >720s	0 <sup>a</sup>	.	.	0	.	.	.
	B: Weiblich	,157	,314	,251	1	,616	-,458	,772
	B: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	B: Alter 17-24	,786	,663	1,404	1	,236	-,514	2,085
	B: Alter 25-34	-,273	,408	,448	1	,503	-1,074	,527
	B: Alter 35-44	,117	,459	,065	1	,799	-,783	1,018
	B: Alter 45-54	-,329	,463	,505	1	,478	-1,237	,579
	B: Alter 55+	0 <sup>a</sup>	.	.	0	.	.	.
	B: Kein Akademiker	-,378	,307	1,524	1	,217	-,979	,222
	B: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	B: Keine Muttersprache Englisch	,305	,453	,453	1	,501	-,583	1,193
	B: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
	B: Erfahrung < 1 Jahr	-1,164	,495	5,533	1	,019	-2,133	-,194
	B: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
	B: < 1 Std. / Woche	-,194	,478	,164	1	,685	-1,130	,743
	B: 1-5 Std. / Woche	,098	,373	,069	1	,793	-,632	,828
	B: 5-15 Std. / Woche	,944	,502	3,538	1	,060	-,040	1,928
	B: > 15 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
	A: Weiblich	-,087	,390	,049	1	,824	-,851	,678
	A: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	A: Alter 17-24	1,167	,980	1,418	1	,234	-,754	3,088
	A: Alter 25-34	2,677	,963	7,729	1	,005	,790	4,564
	A: Alter 35-44	1,392	,632	4,855	1	,028	,154	2,630
	A: Alter 45-54	,120	,541	,049	1	,825	-,940	1,179
	A: Alter 55-64	-,009	,529	,000	1	,987	-1,046	1,028
	A: Alter 65+	0 <sup>a</sup>	.	.	0	.	.	.
	A: Kein Akademiker	-,326	,354	,845	1	,358	-1,019	,368
	A: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	A: Keine Muttersprache Englisch	-,756	,371	4,146	1	,042	-1,483	-,028
	A: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.

	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
A: Erfahrung < 1 Jahr	-2,053	,462	19,758	1	,000	-2,958	-1,148
A: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
A: < 1 Std. / Woche	-,732	1,283	,326	1	,568	-3,246	1,782
A: 1-5 Std. / Woche	-,556	1,018	,298	1	,585	-2,551	1,440
A: 5-15 Std. / Woche	-1,331	1,018	1,710	1	,191	-3,325	,664
A: 15-30 Std. / Woche	-,671	1,084	,383	1	,536	-2,796	1,455
A: >30 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
Link function: Logit.							
a. This parameter is set to zero because it is redundant.							

**Tabelle 9-16: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie einfach der Lösungsvorschlag in die Praxis überführt werden kann (K9). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-7 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet.**

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Thre	Neutral Bewertung	-4,253	1,468	8,390	1	,004	-7,131	-1,375
	Positive Bewertung	-3,211	1,456	4,862	1	,027	-6,065	-,357
Location	Teilstudie 2.1	1,125	,370	9,229	1	,002	,399	1,851
	Teilstudie 2.2	0 <sup>a</sup>	.	.	0	.	.	.
	Wörter: 0-40	-3,870	1,580	5,997	1	,014	-6,967	-,772
	Wörter: 41-80	,325	,450	,522	1	,470	-,557	1,207
	Wörter: 81-120	-,293	,490	,358	1	,549	-1,254	,667
	Wörter: 121+	0 <sup>a</sup>	.	.	0	.	.	.
	Zeit: 0-240s	,552	,974	,321	1	,571	-1,358	2,462
	Zeit: 241-480s	-,263	,479	,302	1	,583	-1,201	,675
	Zeit: 481-720s	-,194	,513	,143	1	,705	-1,200	,812
	Zeit: >720s	0 <sup>a</sup>	.	.	0	.	.	.
	B: Weiblich	-,073	,315	,055	1	,815	-,690	,543
	B: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	B: Alter 17-24	,121	,665	,033	1	,856	-1,183	1,425
	B: Alter 25-34	-,342	,437	,611	1	,435	-1,199	,515
	B: Alter 35-44	-,732	,457	2,561	1	,110	-1,629	,164
	B: Alter 45-54	-,443	,481	,847	1	,357	-1,386	,500
	B: Alter 55+	0 <sup>a</sup>	.	.	0	.	.	.
	B: Kein Akademiker	-,280	,312	,805	1	,370	-,892	,332
	B: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	B: Keine Muttersprache Englisch	-,175	,442	,157	1	,692	-1,042	,691
	B: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
	B: Erfahrung < 1 Jahr	-1,126	,536	4,421	1	,036	-2,177	-,076
	B: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
	B: < 1 Std. / Woche	-,313	,545	,330	1	,566	-1,381	,755
	B: 1-5 Std. / Woche	-,804	,405	3,933	1	,047	-1,599	-,009
	B: 5-15 Std. / Woche	-,098	,511	,036	1	,849	-1,100	,905
	B: > 15 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
	A: Weiblich	-,062	,400	,024	1	,876	-,847	,722
	A: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	A: Alter 17-24	,820	1,102	,554	1	,457	-1,339	2,980
	A: Alter 25-34	1,981	,947	4,373	1	,037	,124	3,838
	A: Alter 35-44	,447	,634	,498	1	,480	-,795	1,689
	A: Alter 45-54	,203	,575	,125	1	,724	-,924	1,331
	A: Alter 55-64	-,203	,536	,143	1	,706	-1,254	,848

	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
A: Alter 65+	0 <sup>a</sup>	.	.	0	.	.	.
A: Kein Akademiker	-,468	,352	1,774	1	,183	-1,158	,221
A: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
A: Keine Muttersprache Englisch	-,535	,368	2,112	1	,146	-1,257	,187
A: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
A: Erfahrung < 1 Jahr	-1,710	,467	13,396	1	,000	-2,626	-,794
A: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
A: < 1 Std. / Woche	1,712	1,643	1,087	1	,297	-1,507	4,932
A: 1-5 Std. / Woche	-,750	1,098	,467	1	,495	-2,901	1,402
A: 5-15 Std. / Woche	-1,058	1,116	,898	1	,343	-3,245	1,130
A: 15-30 Std. / Woche	-,919	1,125	,667	1	,414	-3,124	1,287
A: >30 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
Link function: Logit.							
a. This parameter is set to zero because it is redundant.							

**Tabelle 9-17: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich wie Einfluss der Lösungsvorschlag auf künftige Handlungen des Bewertenden nehmen kann (K10). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-7 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet.**

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Thr	Neutral Bewertung	-3,382	1,195	8,016	1	,005	-5,724	-1,041
	Positive Bewertung	-2,165	1,186	3,334	1	,068	-4,489	,159
Location	Teilstudie 2.1	,561	,286	3,841	1	,050	-6,170E-5	1,123
	Teilstudie 2.2	0 <sup>a</sup>	.	.	0	.	.	.
	Wörter: 0-40	-3,459	1,398	6,118	1	,013	-6,199	-,718
	Wörter: 41-80	-,125	,348	,129	1	,719	-,808	,558
	Wörter: 81-120	-,349	,393	,790	1	,374	-1,120	,421
	Wörter: 121+	0 <sup>a</sup>	.	.	0	.	.	.
	Zeit: 0-240s	-,553	,622	,789	1	,374	-1,772	,667
	Zeit: 241-480s	-,242	,374	,417	1	,518	-,974	,491
	Zeit: 481-720s	-,155	,392	,156	1	,693	-,924	,614
	Zeit: >720s	0 <sup>a</sup>	.	.	0	.	.	.
	B: Weiblich	-,044	,259	,029	1	,865	-,552	,463
	B: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	B: Alter 17-24	-,522	,520	1,008	1	,315	-1,542	,497
	B: Alter 25-34	-,938	,353	7,047	1	,008	-1,631	-,246
	B: Alter 35-44	-1,141	,379	9,059	1	,003	-1,884	-,398
	B: Alter 45-54	-,873	,374	5,453	1	,020	-1,606	-,140
	B: Alter 55+	0 <sup>a</sup>	.	.	0	.	.	.
	B: Kein Akademiker	-,536	,254	4,442	1	,035	-1,034	-,038
	B: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	B: Keine Muttersprache Englisch	-,166	,366	,205	1	,651	-,884	,552
	B: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
	B: Erfahrung < 1 Jahr	-,105	,439	,057	1	,811	-,966	,756
	B: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
	B: < 1 Std. / Woche	-,698	,412	2,869	1	,090	-1,505	,110
	B: 1-5 Std. / Woche	-,235	,319	,540	1	,462	-,860	,391
	B: 5-15 Std. / Woche	-,315	,380	,689	1	,406	-1,060	,429
	B: > 15 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
	A: Weiblich	-,252	,322	,611	1	,434	-,884	,380
	A: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	A: Alter 17-24	,354	,858	,170	1	,680	-1,328	2,036
	A: Alter 25-34	,750	,646	1,349	1	,245	-,516	2,017
	A: Alter 35-44	-,238	,501	,225	1	,635	-1,220	,745
	A: Alter 45-54	-,819	,458	3,196	1	,074	-1,717	,079
	A: Alter 55-64	-,617	,461	1,788	1	,181	-1,520	,287
	A: Alter 65+	0 <sup>a</sup>	.	.	0	.	.	.
	A: Kein Akademiker	-,259	,291	,791	1	,374	-,829	,311
	A: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	A: Keine Muttersprache Englisch	,001	,303	,000	1	,997	-,593	,595



	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
A: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
A: Erfahrung < 1 Jahr	-,708	,378	3,505	1	,061	-1,449	,033
A: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
A: < 1 Std. / Woche	,684	1,150	,353	1	,552	-1,571	2,938
A: 1-5 Std. / Woche	,354	,907	,153	1	,696	-1,422	2,131
A: 5-15 Std. / Woche	-,461	,882	,274	1	,601	-2,189	1,266
A: 15-30 Std. / Woche	-,499	,911	,300	1	,584	-2,285	1,286
A: >30 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
Link function: Logit.							
a. This parameter is set to zero because it is redundant.							

**Tabelle 9-18: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich zu welchem Grad der Bewertende den Lösungsvorschlag weiterempfehlen würde (K11). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-5 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet.**

		Estimate	Std. Error	Wald	df	Sig.	95% Conf. Interval	
							Lower Bound	Upper Bound
Thr	Neutral Bewertung	-3,590	1,371	6,855	1	,009	-6,277	-,903
	Positive Bewertung	-2,735	1,364	4,019	1	,045	-5,410	-,061
Location	Teilstudie 2.1	,802	,317	6,407	1	,011	,181	1,424
	Teilstudie 2.2	0 <sup>a</sup>	.	.	0	.	.	.
	Wörter: 0-40	-4,575	1,791	6,522	1	,011	-8,086	-1,064
	Wörter: 41-80	,018	,388	,002	1	,962	-,743	,779
	Wörter: 81-120	,129	,429	,090	1	,764	-,712	,970
	Wörter: 121+	0 <sup>a</sup>	.	.	0	.	.	.
	Zeit: 0-240s	,062	,745	,007	1	,934	-1,399	1,522
	Zeit: 241-480s	-,362	,412	,774	1	,379	-1,169	,445
	Zeit: 481-720s	,076	,437	,030	1	,863	-,782	,933
	Zeit: >720s	0 <sup>a</sup>	.	.	0	.	.	.
	B: Weiblich	-,351	,288	1,484	1	,223	-,916	,214
	B: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	B: Alter 17-24	,183	,601	,092	1	,761	-,996	1,361
	B: Alter 25-34	-,263	,388	,459	1	,498	-1,022	,497
	B: Alter 35-44	-1,308	,406	10,397	1	,001	-2,103	-,513
	B: Alter 45-54	-,737	,410	3,239	1	,072	-1,540	,066
	B: Alter 55+	0 <sup>a</sup>	.	.	0	.	.	.
	B: Kein Akademiker	-,077	,276	,078	1	,779	-,617	,463
	B: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	B: Keine Muttersprache Englisch	-,377	,401	,885	1	,347	-1,164	,409
	B: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
	B: Erfahrung < 1 Jahr	-,721	,483	2,231	1	,135	-1,668	,225
	B: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
	B: < 1 Std. / Woche	-,080	,463	,030	1	,862	-,988	,827
	B: 1-5 Std. / Woche	-,053	,354	,022	1	,881	-,747	,641
	B: 5-15 Std. / Woche	-,165	,417	,156	1	,693	-,982	,652
	B: > 15 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
	A: Weiblich	-,696	,370	3,540	1	,060	-1,421	,029
	A: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	A: Alter 17-24	1,142	1,035	1,219	1	,270	-,886	3,171
	A: Alter 25-34	1,788	,761	5,528	1	,019	,298	3,279
	A: Alter 35-44	,157	,553	,080	1	,777	-,927	1,240
	A: Alter 45-54	-,291	,508	,328	1	,567	-1,287	,705
	A: Alter 55-64	-,476	,496	,920	1	,338	-1,448	,497
	A: Alter 65+	0 <sup>a</sup>	.	.	0	.	.	.
	A: Kein Akademiker	-,457	,321	2,030	1	,154	-1,085	,172
	A: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	A: Keine Muttersprache Englisch	-,106	,341	,096	1	,756	-,774	,563

	Estimate	Std. Error	Wald	df	Sig.	95% Conf. Interval	
						Lower Bound	Upper Bound
A: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
A: Erfahrung < 1 Jahr	-1,951	,435	20,091	1	,000	-2,804	-1,098
A: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
A: < 1 Std. / Woche	,884	1,335	,438	1	,508	-1,733	3,501
A: 1-5 Std. / Woche	-,282	1,072	,069	1	,793	-2,382	1,819
A: 5-15 Std. / Woche	-,920	1,048	,770	1	,380	-2,975	1,135
A: 15-30 Std. / Woche	-1,533	1,069	2,057	1	,152	-3,628	,562
A: >30 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
Link function: Logit.							
a. This parameter is set to zero because it is redundant.							

**Tabelle 9-19: Ordinales Regressionsmodell für die Bewertungskategorie hinsichtlich zu welchem Grad der Bewertende den Lösungsvorschlag selbst nutzen würde (K12). Der Test für die Validität des Regressionsmodells auf Basis der 3er-Likert-Skala ist in Tabelle 9-5 abgebildet. Der Referenzwert für den Vergleich ist eine negative Bewertung hinsichtlich der Bewertungskategorie. Die demografischen Daten sind mit ‚B‘ für Bewertende und ‚A‘ für Autoren gekennzeichnet.**

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Thr	Neutral Bewertung	-2,829	1,217	5,401	1	,020	-5,215	-,443
	Positive Bewertung	-1,605	1,210	1,760	1	,185	-3,977	,767
Location	Teilstudie 2.1	,668	,292	5,245	1	,022	,096	1,240
	Teilstudie 2.2	0 <sup>a</sup>	.	.	0	.	.	.
	Wörter: 0-40	-3,670	1,422	6,665	1	,010	-6,456	-,884
	Wörter: 41-80	-,008	,357	,001	1	,982	-,707	,691
	Wörter: 81-120	-,266	,402	,438	1	,508	-1,055	,522
	Wörter: 121+	0 <sup>a</sup>	.	.	0	.	.	.
	Zeit: 0-240s	,404	,680	,353	1	,553	-,930	1,738
	Zeit: 241-480s	-,010	,385	,001	1	,979	-,764	,744
	Zeit: 481-720s	,033	,401	,007	1	,934	-,752	,818
	Zeit: >720s	0 <sup>a</sup>	.	.	0	.	.	.
	B: Weiblich	-,199	,269	,545	1	,460	-,726	,329
	B: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	B: Alter 17-24	,366	,548	,445	1	,505	-,708	1,439
	B: Alter 25-34	,052	,358	,021	1	,884	-,649	,753
	B: Alter 35-44	-,649	,374	3,015	1	,082	-1,381	,084
	B: Alter 45-54	-,529	,372	2,029	1	,154	-1,258	,199
	B: Alter 55+	0 <sup>a</sup>	.	.	0	.	.	.
	B: Kein Akademiker	-,074	,260	,082	1	,774	-,583	,434
	B: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	B: Keine Muttersprache Englisch	-,376	,377	,993	1	,319	-1,114	,363
	B: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.
	B: Erfahrung < 1 Jahr	-,648	,451	2,062	1	,151	-1,532	,236
	B: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
	B: < 1 Std. / Woche	-,255	,431	,352	1	,553	-1,099	,588
	B: 1-5 Std. / Woche	-,221	,329	,452	1	,501	-,867	,424
	B: 5-15 Std. / Woche	-,537	,385	1,947	1	,163	-1,292	,217
	B: > 15 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
	A: Weiblich	-,303	,333	,828	1	,363	-,956	,350
	A: Männlich	0 <sup>a</sup>	.	.	0	.	.	.
	A: Alter 17-24	,854	,914	,873	1	,350	-,937	2,644
	A: Alter 25-34	1,637	,679	5,823	1	,016	,308	2,967
	A: Alter 35-44	,263	,505	,272	1	,602	-,726	1,252
	A: Alter 45-54	-,244	,457	,284	1	,594	-1,140	,653
	A: Alter 55-64	-,064	,454	,020	1	,888	-,953	,825
	A: Alter 65+	0 <sup>a</sup>	.	.	0	.	.	.
	A: Kein Akademiker	-,559	,298	3,521	1	,061	-1,142	,025
	A: Akademiker	0 <sup>a</sup>	.	.	0	.	.	.
	A: Keine Muttersprache Englisch	,092	,314	,086	1	,770	-,523	,706
	A: Muttersprache Englisch	0 <sup>a</sup>	.	.	0	.	.	.

	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
A: Erfahrung < 1 Jahr	-1,464	,398	13,541	1	,000	-2,244	-,684
A: Erfahrung > 1 Jahr	0 <sup>a</sup>	.	.	0	.	.	.
A: < 1 Std. / Woche	,132	1,164	,013	1	,910	-2,150	2,413
A: 1-5 Std. / Woche	-,085	,933	,008	1	,928	-1,914	1,744
A: 5-15 Std. / Woche	-,695	,918	,573	1	,449	-2,495	1,104
A: 15-30 Std. / Woche	-1,108	,940	1,390	1	,238	-2,951	,734
A: >30 Std. / Woche	0 <sup>a</sup>	.	.	0	.	.	.
Link function: Logit.							
a. This parameter is set to zero because it is redundant.							



## Literaturverzeichnis

- von Ahn, Luis und Laura Dabbish. 2008. Designing Games with a Purpose. *Commun. ACM* 51, Nr. 8 (August): 58–67. doi:10.1145/1378704.1378719.
- Alexiou, Aikaterini und Fotini Paraskeva. 2010. Enhancing self-regulated learning skills through the implementation of an e-portfolio tool. *Procedia - Social and Behavioral Sciences* 2, Nr. 2. Innovation and Creativity in Education: 3048–3054. doi:10.1016/j.sbspro.2010.03.463.
- Amabile, Teresa M. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology* 43, Nr. 5: 997–1013. doi:10.1037/0022-3514.43.5.997.
- Amir, Ofra, David G. Rand und Ya'akov Kobi Gal. 2012. Economic Games on the Internet: The Effect of \$1 Stakes. *PLOS ONE* 7, Nr. 2 (21. Februar): e31461. doi:10.1371/journal.pone.0031461.
- Anzalone, Filippa Marullo. 2010. Education for the Law: Reflective Education for the Law. In: *Handbook of Reflection and Reflective Inquiry: Mapping a Way of Knowing for Professional Reflective Inquiry*, hg. von Nona Lyons, 85–99. Boston, MA: Springer US. doi:10.1007/978-0-387-85744-2\_5, (zugegriffen: 16. Juni 2019).
- Ardaiz-Villanueva, Oscar, Xabier Nicuesa-Chacón, Oscar Brene-Artazcoz, María Luisa Sanz de Acedo Lizarraga und María Teresa Sanz de Acedo Baquedano. 2011. Evaluation of computer tools for idea generation and team formation in project-based learning. *Computers & Education* 56, Nr. 3 (1. April): 700–711. doi:10.1016/j.compedu.2010.10.012.
- Ardichvili, Alexander, Vaughn Page und Tim Wentling. 2003. Motivation and barriers to participation in virtual knowledge-sharing communities of practice. *Journal of Knowledge Management* 7, Nr. 1 (1. März): 64–77. doi:10.1108/13673270310463626.
- Ardichvili, Alexandre, Martin Maurer, Wei Li, Tim Wentling und Reed Stuedemann. 2006. Cultural influences on knowledge sharing through online communities of practice. *Journal of Knowledge Management* 10, Nr. 1 (1. Januar): 94–107. doi:10.1108/13673270610650139.
- Argyris, C. und D.A. Schön. 1996. *Organizational learning II: Theory, method, and practice*. Addison-Wesley.
- Atkins, Sue und Kathy Murphy. 1993. Reflection: a review of the literature. *Journal of Advanced Nursing* 18, Nr. 8 (1. August): 1188–1192. doi:10.1046/j.1365-2648.1993.18081188.x.
- Bandura, Albert. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84, Nr. 2: 191–215. doi:10.1037/0033-295X.84.2.191.
- . 1993. Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist* 28, Nr. 2: 117–148. doi:10.1207/s15326985ep2802\_3.
- Bannert, Maria. 2006. Effects of Reflection Prompts When Learning with Hypermedia. *Journal of Educational Computing Research* 35, Nr. 4 (1. Dezember): 359–375. doi:10.2190/94V6-R58H-3367-G388.
- . 2007. *Metakognition beim Lernen mit Hypermedien*. Waxmann Verlag.
- . 2009. Promoting Self-Regulated Learning Through Prompts. *Zeitschrift für Pädagogische Psychologie* 23, Nr. 2 (1. Januar): 139–145. doi:10.1024/1010-0652.23.2.139.
- Baumer, Eric PS. 2015. Reflective Informatics: Conceptual Dimensions for Designing Technologies of Reflection. In: *Proceedings of CHI 2015*.

- Baumer, Eric P.S., Vera Khovanskaya, Mark Matthews, Lindsay Reynolds, Victoria Schwanda Sosik und Geri Gay. 2014. Reviewing Reflection: On the Use of Reflection in Interactive System Design. In: *Proceedings of the 2014 Conference on Designing Interactive Systems*, 93–102. DIS '14. New York, NY, USA: ACM. doi:10.1145/2598510.2598598.
- Behrend, Tara S., David J. Sharek, Adam W. Meade und Eric N. Wiebe. 2011. The viability of crowdsourcing for survey research. *Behavior Research Methods* 43, Nr. 3 (25. März): 800–813. doi:10.3758/s13428-011-0081-0.
- Bell, Philip. 2000. Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education* 22, Nr. 8 (1. August): 797–817. doi:10.1080/095006900412284.
- Bernstein, Michael S., Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell und Katrina Panovich. 2015. Soylent: A Word Processor with a Crowd Inside. *Commun. ACM* 58, Nr. 8 (Juli): 85–94. doi:10.1145/2791285.
- Berthold, Kirsten, Matthias Nückles und Alexander Renkl. 2007. Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction* 17, Nr. 5 (Oktober): 564–577. doi:10.1016/j.learninstruc.2007.09.007.
- Bimrose, Jenny, Alan Brown, Teresa Holocher-Ertl, Barbara Kieslinger, Christine Kunzmann, Michael Prilla, Andreas Schmidt und Carmen Wolf. 2014a. Introducing learning innovation in public employment services. What role can facilitation play? In: *Proceedings of International Conference on E-Learning at the Workplace (ICELW) 2014, New York City, USA, June 11-13, 2014*. [http://publications.andreas.schmidt.name/ICELW2014\\_EmployID.pdf](http://publications.andreas.schmidt.name/ICELW2014_EmployID.pdf).
- . 2014b. The Role of Facilitation in Technology-Enhanced Learning for Public Employment Services. *International Journal of Advanced Corporate Learning (iJAC)* 7, Nr. 3 (14. Oktober): 56–63.
- Blohm, Ivo, Ulrich Bretschneider, Michael Huber, Jan Marco Leimeister und Helmut Krcmar. 2009. Collaborative Filtering in Ideenwettbewerben - Evaluation zweier Skalen zur Teilnehmer-Bewertung in Ideenwettbewerben. In: *Virtuelle Organisation und Neue Medien 2009*, hg. von Klaus Meißner und Martin Engelen, 365–377. Dresden: TUDpress, 1. Oktober. <https://www.alexandria.unisg.ch/220523/>.
- Blohm, Ivo, Ulrich Bretschneider, Jan Marco Leimeister und Helmut Krcmar. 2010. Does Collaboration among Participants Lead to Better Ideas in IT-based Idea Competitions? An Empirical Investigation. In: *Proceedings of the 43rd Hawaii International Conference on System Sciences 2010 (HICSS-43)*, 1–10. Los Alamitos: IEEE Computer Society, 1. Januar. <https://www.alexandria.unisg.ch/220669/>.
- Blohm, Ivo, Christoph Riedl, Johann Füller und Jan Marco Leimeister. 2016. Rate or Trade? Identifying Winning Ideas in Open Idea Sourcing. *Information Systems Research* 27, Nr. 1 (1. März): 27–48. doi:10.1287/isre.2015.0605.
- Blunk, Oliver und Michael Prilla. 2015. Prompting users to facilitate support needs in collaborative reflection. In: *Proceedings of the 5th International Workshop on Awareness and Reflection in Technology Enhanced Learning (ARTEL 2015) in conjunction with the ECTEL 2015 conference*, hg. von M. Kravcik, A. Mikroyannidis, V. Pammer, M. Prilla, und T. D. Ullmann, 1465:43–57. CEUR-WS. <http://ceur-ws.org/Vol-1465/>.



- . 2017a. Supporting Communities of Practice in Public Administrations: Factors Influencing Adoption and Readiness. In: *Proceedings of the 8th International Conference on Communities and Technologies*, 36–45. C&T '17. New York, NY, USA: ACM. doi:10.1145/3083671.3083699.
- . 2017b. Prompting to Support Reflection: A Workplace Study. In: *Data Driven Approaches in Digital Education*, 367–372. Lecture Notes in Computer Science. Springer, Cham, 12. September. doi:10.1007/978-3-319-66610-5\_28, [https://link.springer.com/chapter/10.1007/978-3-319-66610-5\\_28](https://link.springer.com/chapter/10.1007/978-3-319-66610-5_28) (zugegriffen: 14. September 2017).
- Boerboom, Tobias B. B., Debbie Jaarsma, Diana H. J. M. Dolmans, Albert J. J. A. Scherpbier, Nicole J. J. M. Mastenbroek und Peter Van Beukelen. 2011. Peer group reflection helps clinical teachers to critically reflect on their teaching. *Medical Teacher* 33, Nr. 11 (1. November): e615–e623. doi:10.3109/0142159X.2011.610840.
- Bolton, Gillie. 2010. *Reflective Practice: Writing and Professional Development*. SAGE.
- Bonney, Rick, Tina B. Phillips, Heidi L. Ballard und Jody W. Enck. 2016. Can citizen science enhance public understanding of science? *Public Understanding of Science* 25, Nr. 1 (1. Januar): 2–16. doi:10.1177/0963662515607406.
- van den Boom, Gerard, Fred Paas, Jeroen J. G. van Merriënboer und Tamara van Gog. 2004. Reflection prompts and tutor feedback in a web-based learning environment: effects on students' self-regulated learning competence. *Computers in Human Behavior* 20, Nr. 4 (Juli): 551–567. doi:10.1016/j.chb.2003.10.001.
- van den Boom, Gerard, Fred Paas und Jeroen JG van Merriënboer. 2007. Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction* 17, Nr. 5: 532–548.
- Borzillo, Stefano, Stéphane Aznar und Achim Schmitt. 2011. A journey through communities of practice: How and why members move from the periphery to the core. *European Management Journal* 29, Nr. 1 (Februar): 25–42. doi:10.1016/j.emj.2010.08.004.
- Boud, David, Peter Cressey und Peter Docherty. 2006. *Productive Reflection at Work: Learning for Changing Organizations*. Routledge.
- Boud, David, Rosemary Keogh und David Walker. 1985. *Reflection: Turning Experience Into Learning*. Routledge.
- Boyd, Evelyn M. und Ann W. Fales. 1983. Reflective Learning: Key to Learning from Experience. *Journal of Humanistic Psychology* 23, Nr. 2 (1. April): 99–117. doi:10.1177/0022167883232011.
- Brabham, Daren C. 2008. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence* 14, Nr. 1 (1. Februar): 75–90. doi:10.1177/1354856507084420.
- . 2009. Crowdsourcing the Public Participation Process for Planning Projects. *Planning Theory* 8, Nr. 3 (1. August): 242–262. doi:10.1177/1473095209104824.
- Brookfield, Stephen. 2010. Critical Reflection as an Adult Learning Process. In: *Handbook of Reflection and Reflective Inquiry: Mapping a Way of Knowing for Professional Reflective Inquiry*, hg. von Nona Lyons, 215–236. Boston, MA: Springer US. doi:10.1007/978-0-387-85744-2\_11, (zugegriffen: 16. Juni 2019).
- Brookfield, Stephen D. 1987. *Developing critical thinkers: Challenging adults to explore alternative ways of thinking and acting*. Developing critical thinkers: Challenging adults to explore alternative ways of thinking and acting. San Francisco, CA, US: Jossey-Bass.

- Brown, Alan. 1997. A dynamic model of occupational identity formation. In: *Promoting vocational education and training : European perspectives*, hg. von Alan Brown, 59–67. Hämeenlinna: Tampereen yliopiston opettajankoulutuslaitos. <http://wrap.warwick.ac.uk/57844/> (zugegriffen: 30. Juni 2019).
- Brown, Alan, Jenny Bimrose, Sally-Anne Barnes und Deirdre Hughes. 2012. The role of career adaptabilities for mid-career changers. *Journal of Vocational Behavior* 80, Nr. 3. Career Adaptability (Juni): 754–761. doi:10.1016/j.jvb.2012.01.003.
- Brown, Alan, Jenny Bimrose, Teresa Holocher-Ertl, Barbara Kieslinger, Urša Dolinar, Marina Hudolin, Anette Scopetta und Fiona Smith. 2015. D2.1 First Activity Report for Identity Transformation in Public Employment Services (PES). Hg. von Teresa Holocher-Ertl und Jenny Bimrose. 13. Februar. interne Version.
- Brown, John Seely und Paul Duguid. 1991. Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science* 2, Nr. 1: 40–57.
- Buhrmester, Michael, Tracy Kwang und Samuel D. Gosling. 2011. Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science* 6, Nr. 1 (1. Januar): 3–5. doi:10.1177/1745691610393980.
- Butler, Deborah L. und Philip H. Winne. 1995. Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research* 65, Nr. 3 (1. September): 245–281. doi:10.3102/00346543065003245.
- Callahan, Shawn. 2004. Cultivating a Public Sector Knowledge Management Community of Practice. <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-59140-200-8.ch022>: 267–281. doi:10.4018/978-1-59140-200-8.ch022.
- Carroll, John M., Dennis C. Neale, Philip L. Isenhour, Mary Beth Rosson und D. Scott Mccrickard. 2003. Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human–Computer Studies*: 605–632. doi:10.1016/S1071-5819(03)00024-7.
- Cheetham, Graham und Geoff Chivers. 2000. A new look at competent professional practice. *Journal of European Industrial Training* 24, Nr. 7 (1. Oktober): 374–383. doi:10.1108/03090590010349827.
- Chen, Nian-Shing, Chun-Wang Wei, Kuen-Ting Wu und Lorna Uden. 2009. Effects of high level prompts and peer assessment on online learners’ reflection levels. *Computers & Education* 52, Nr. 2 (Februar): 283–291. doi:10.1016/j.compedu.2008.08.007.
- Chirema, Kathleen D. 2007. The use of reflective journals in the promotion of reflection and learning in post-registration nursing students. *Nurse Education Today* 27, Nr. 3 (1. April): 192–202. doi:10.1016/j.nedt.2006.04.007.
- Choe, Eun Kyoung, Bongshin Lee, Haining Zhu, Nathalie Henry Riche und Dominikus Baur. 2017. Understanding Self-Reflection: How People Reflect on Personal Data Through Visual Data Exploration. In: . <https://www.microsoft.com/en-us/research/publication/understanding-self-reflection-people-reflect/>.
- Choi, Ikseon, Susan M. Land und Alfred Turgeon. 2008. Instructor Modeling and Online Question Prompts for Supporting Peer-Questioning during Online Discussion. *Journal of Educational Technology Systems* 36, Nr. 3 (1. März): 255–275. doi:10.2190/ET.36.3.c.
- Christensen, Kasper, Joachim Scholderer, Stine Alm Hersleth, Tormod Næs, Knut Kvaal, Torulf Mollestad, Nina Veflen und Einar Risvik. 2018. How good are ideas

- identified by an automatic idea detection system? *Creativity and Innovation Management* 27, Nr. 1: 23–31. doi:10.1111/caim.12260.
- Christiaans, Henri H. C. M. 2002. Creativity as a Design Criterion. *Creativity Research Journal* 14, Nr. 1 (Januar): 41–54. doi:10.1207/S15326934CRJ1401\_4.
- Chung, Phillip H und Michael D Byrne. 2004. Visual cues to reduce errors in a routine procedural task. In: *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*.
- Chung, Phillip H. und Michael D. Byrne. 2008. Cue effectiveness in mitigating postcompletion errors in a routine procedural task. *International Journal of Human-Computer Studies* 66, Nr. 4 (April): 217–232. doi:10.1016/j.ijhcs.2007.09.001.
- Clarà, Marc. 2015. What Is Reflection? Looking for Clarity in an Ambiguous Notion. *Journal of Teacher Education* 66, Nr. 3 (1. Mai): 261–271. doi:10.1177/0022487114552028.
- Clardy, Alan. 2018. 70-20-10 and the Dominance of Informal Learning: A Fact in Search of Evidence. *Human Resource Development Review* 17, Nr. 2 (1. Juni): 153–178. doi:10.1177/1534484318759399.
- Clarke, Maggie. 2004. Reflection: Journals and Reflective Questions: a Strategy for Professional Learning. *Australian Journal of Teacher Education* 29, Nr. 2 (1. November). doi:10.14221/ajte.2004v29n2.2, <http://ro.ecu.edu.au/ajte/vol29/iss2/2>.
- Clouder, Lynn. 2000. Reflective Practice in Physiotherapy Education: A critical conversation. *Studies in Higher Education* 25, Nr. 2 (1. Juni): 211–223. doi:10.1080/713696142.
- Cobb, Paul, Ada Boufi, Kay McClain und Joy Whitenack. 1997. Reflective Discourse and Collective Reflection. *Journal for Research in Mathematics Education* 28, Nr. 3: 258–277. doi:10.2307/749781.
- Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates.
- Cole, Forrester, Kevin Sanik, Doug DeCarlo, Adam Finkelstein, Thomas Funkhouser, Szymon Rusinkiewicz und Manish Singh. 2009. How Well Do Line Drawings Depict Shape? In: *ACM SIGGRAPH 2009 Papers*, 28:1–28:9. SIGGRAPH '09. New York, NY, USA: ACM. doi:10.1145/1576246.1531334, (zugegriffen: 27. Mai 2019).
- Craig, Cheryl J. 2010. Reflective Practice in the Professions: Teaching. In: *Handbook of Reflection and Reflective Inquiry: Mapping a Way of Knowing for Professional Reflective Inquiry*, hg. von Nona Lyons, 189–214. Boston, MA: Springer US. doi:10.1007/978-0-387-85744-2\_10, [https://doi.org/10.1007/978-0-387-85744-2\\_10](https://doi.org/10.1007/978-0-387-85744-2_10) (zugegriffen: 16. Juni 2019).
- Cressey, P., D. Boud und P. Docherty. 2006. The emergence of productive reflection. *Productive reflection at work: learning for changing organizations* (1. Januar): 11–26. doi:10.4324/9780203001745.
- Crump, Matthew J. C., John V. McDonnell und Todd M. Gureckis. 2013. Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE* 8, Nr. 3 (13. März): e57410. doi:10.1371/journal.pone.0057410.
- Daudelin, Marilyn Wood. 1996. Learning from experience through reflection. *Organizational Dynamics* 24, Nr. 3: 36–48. doi:10.1016/S0090-2616(96)90004-2.

- Davis, Elizabeth A. 2003. Prompting Middle School Science Students for Productive Reflection: Generic and Directed Prompts. *Journal of the Learning Sciences* 12, Nr. 1 (1. Januar): 91–142. doi:10.1207/S15327809JLS1201\_4.
- . 2006. Characterizing productive reflection among preservice elementary teachers: Seeing what matters. *Teaching and Teacher Education* 22, Nr. 3 (1. April): 281–301. doi:10.1016/j.tate.2005.11.005.
- Davis, Elizabeth A. und Marcia C. Linn. 2000. Scaffolding students' knowledge integration: prompts for reflection in KIE. *International Journal of Science Education* 22, Nr. 8 (1. August): 819–837. doi:10.1080/095006900412293.
- Dawes, John. 2008. Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-Point, 7-Point and 10-Point Scales. *International Journal of Market Research* 50, Nr. 1 (1. Januar): 61–104. doi:10.1177/147078530805000106.
- D'Cruz, Heather, Philip Gillingham und Sebastian Melendez. 2007. Reflexivity, its Meanings and Relevance for Social Work: A Critical Review of the Literature. *The British Journal of Social Work* 37, Nr. 1 (1. Januar): 73–90. doi:10.1093/bjsw/bcl001.
- Dean, Douglas L, Jillian M Hender, Thomas L Rodgers und Eric L Santanen. 2006. Identifying Quality, Novel, and Creative Ideas: Constructs and Scales for Idea Evaluation: 56.
- Deitz, Samuel M. und Leslie W. Malone. 1985. Stimulus control terminology. *The Behavior Analyst* 8, Nr. 2: 259–264.
- Denham, Bryan E. 2017. *Categorical statistics for communication research*. <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1441133> (zugegriffen: 28. August 2019).
- Dewey, John. 1933. *How we think: A restatement of the relation of reflective thinking to the educative process*. London ; New York: DC Heath.
- Difallah, Djellel, Gianluca Demartini und Philippe Cudré-Mauroux. 2012. Mechanical Cheat: Spamming Schemes and Adversarial Techniques on Crowdsourcing Platforms. *CrowdSearch 2012 workshop at WWW*: 26–30.
- Dillenbourg, Pierre. 1999. What do you mean by collaborative learning? <https://telearn.archives-ouvertes.fr/hal-00190240> (zugegriffen: 10. Juni 2019).
- Dimitrova, Vania, Antonija Mitrovic, Alicja Piotrkowicz, Lydia Lau und Amali Weerasinghe. 2017. Using Learning Analytics to Devise Interactive Personalised Nudges for Active Video Watching. In: *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 22–31. UMAP '17. New York, NY, USA: ACM. doi:10.1145/3079628.3079683, <http://doi.acm.org/10.1145/3079628.3079683> (zugegriffen: 15. Mai 2019).
- Doan, AnHai und Raghu Ramakrishnan. 2014. The practice of crowdsourcing is transforming the Web and giving rise to a new field. In: .
- Dyke, Martin. 2006. The role of the 'Other' in reflection, knowledge formation and action in a late modernity. *International Journal of Lifelong Education* 25, Nr. 2 (1. März): 105–123. doi:10.1080/02601370500510728.
- Egelman, Serge, Ed H Chi und Steven Dow. 2014. Crowdsourcing in HCI Research. In: *Ways of Knowing in HCI*, 267–289. Springer.
- Elango, S., R. C. Jutti und L. K. Lee. 2005. Portfolio as a learning tool: students' perspective. *Annals of the Academy of Medicine, Singapore* 34, Nr. 8: 511–4.
- Ellena, Thierry, Aleksandar Subic, Helmy Mustafa und Toh Yen Pang. 2016. The Helmet Fit Index – An intelligent tool for fit assessment and design customisation. *Applied Ergonomics* 55 (Juli): 194–207. doi:10.1016/j.apergo.2016.02.008.

- EmployID. 2016. *Empowering Change in Public Employment Services: The EmployID Approach*. Hg. von Andreas Schmidt und Christine Kunzmann. <https://employid.eu>.
- . 2018. *EmployID Deliverables Year 4 D[2-9].4*. Hg. von Andreas Schmidt und Christine Kunzmann. (interne Version).
- Ennis, Robert H. 1993. Critical thinking assessment. *Theory Into Practice* 32, Nr. 3 (1. Juni): 179–186. doi:10.1080/00405849309543594.
- Entwistle, Noel und Paul Ramsden. 1983. *Understanding Student Learning*. Routledge. doi:10.4324/9781315718637, <https://www.taylorfrancis.com/books/9781315718637> (zugegriffen: 8. Juni 2019).
- Eraut, Michael. 2000. Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* 70, Nr. 1: 113–136. doi:10.1348/000709900158001.
- . 2004. Informal learning in the workplace. *Studies in Continuing Education* 26, Nr. 2: 247–273. doi:10.1080/158037042000225245.
- Fails, Jerry Alan und Dan R. Olsen Jr. 2003. Interactive Machine Learning. In: *Proceedings of the 8th International Conference on Intelligent User Interfaces*, 39–45. IUI '03. New York, NY, USA: ACM. doi:10.1145/604045.604056, (zugegriffen: 24. Mai 2019).
- Faltin, Nils, Simon Schwantzer und Margret Jung. 2014. Activity recommendation app - software to evaluate the usefulness of improvement recommendations created in a team. In: *Proceedings of the 4th Workshop on Awareness and Reflection in Technology-Enhanced Learning. In conjunction with the 9th European Conference on Technology Enhanced Learning: Open Learning and Teaching in Educational Communities (ECTEL 2014)*, hg. von Milos Kravcik, Alexander Mikroyannidis, Viktoria Pammer, Michael Prilla, Thomas Daniel Ullmann, und Fridolin Wild, 77–79. Graz, Austria, 15. September.
- Fast, Ethan, Binbin Chen und Michael S. Bernstein. 2016. Empath: Understanding Topic Signals in Large-Scale Text. In: , 4647–4657. ACM. doi:10.1145/2858036.2858535.
- Ferrario, Maria Angela, Will Simm, Adrian Gradinar, Stephen Forshaw, Marcia Tavares Smith, Thomas Lee, Ian Smith und Jon Whittle. 2017. Computing and Mental Health: Intentionality and Reflection at the Click of a Button. In: *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 1–10. PervasiveHealth '17. New York, NY, USA: ACM. doi:10.1145/3154862.3154877, (zugegriffen: 29. Mai 2019).
- Fessl, Angela, Oliver Blunk, Michael Prilla und Viktoria Pammer. 2017. The known Universe of Reflection Guidance: A Literature Review. *International Journal of Technology Enhanced Learning* 9, Nr. 2–3 (1. Januar): 103–125. doi:10.1504/IJTEL.2017.084491.
- Fessl, Angela, Gudrun Wesiak, Verónica Rivera-Pelayo, Sandra Feyertag und Viktoria Pammer. 2015. In-App Reflection Guidance for Workplace Learning. In: *Design for Teaching and Learning in a Networked World*, hg. von Gráinne Conole, Tomaž Klobučar, Christoph Rensing, Johannes Konert, und Elise Lavoué, 85–99. Lecture Notes in Computer Science. Springer International Publishing.
- Field, Andy P. 2009. *Discovering statistics using SPSS: and sex, drugs and rock „n“ roll*. 3rd ed. Los Angeles: SAGE Publications.

- Finlay, Linda. 2002. "Outing" the Researcher: The Provenance, Process, and Practice of Reflexivity. *Qualitative Health Research* 12, Nr. 4 (1. April): 531–545. doi:10.1177/104973202129120052.
- Fiorella, Logan und Richard E. Mayer. 2012. Paper-Based Aids for Learning with a Computer-Based Game. *Journal of Educational Psychology* 104, Nr. 4 (November): 1074–1082.
- Fivush, Robyn, Jacquelyn T. Gray und Fayne A. Fromhoff. 1987. Two-year-old talk about the past. *Cognitive Development* 2, Nr. 4 (Oktober): 393–409. doi:10.1016/S0885-2014(87)80015-1.
- Fleck, Rowanne und Geraldine Fitzpatrick. 2006. Supporting collaborative reflection with passive image capture. In: . Carry-le-Rouet, France, Mai. [http://coop-2012.grenoble-inp.fr/conference\\_serie.html](http://coop-2012.grenoble-inp.fr/conference_serie.html) (zugegriffen: 10. Juni 2019).
- . 2009. Teachers' and tutors' social reflection around SenseCam images. *International Journal of Human-Computer Studies* 67, Nr. 12. Collocated Social Practices Surrounding Photos (Dezember): 1024–1036. doi:10.1016/j.ijhcs.2009.09.004.
- . 2010. Reflecting on Reflection: Framing a Design Landscape. In: *Proceedings of the 22Nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, 216–223. OZCHI '10. New York, NY, USA: ACM. doi:10.1145/1952222.1952269, <http://doi.acm.org/10.1145/1952222.1952269> (zugegriffen: 19. September 2014).
- Fost, Dan. 2008. Despite Silicon Valley Optimism, a Disease Resists Cure. *The New York Times*, 14. April, Abschn. Technology. <https://www.nytimes.com/2008/04/14/technology/14homer.html> (zugegriffen: 29. Mai 2019).
- Freshwater, Dawn und Gary Rolfe. 2001. Critical reflexivity: A politically and ethically engaged research method for nursing. *NT Research* 6, Nr. 1 (1. Januar): 526–537. doi:10.1177/136140960100600109.
- Furberg, A. 2009. Socio-cultural aspects of prompting student reflection in Web-based inquiry learning environments. *Journal of Computer Assisted Learning* 25, Nr. 4 (1. August): 397–409. doi:10.1111/j.1365-2729.2009.00320.x.
- Gadiraju, Ujwal, Besnik Fetahu und Ricardo Kawase. 2015. Training Workers for Improving Performance in Crowdsourcing Microtasks. In: *Design for Teaching and Learning in a Networked World*, hg. von Gránne Conole, Tomaž Klobučar, Christoph Rensing, Johannes Konert, und Élise Lavoué, 100–114. Lecture Notes in Computer Science 9307. Springer International Publishing. [http://link.springer.com/chapter/10.1007/978-3-319-24258-3\\_8](http://link.springer.com/chapter/10.1007/978-3-319-24258-3_8) (zugegriffen: 21. September 2015).
- Gadiraju, Ujwal, Ricardo Kawase und Stefan Dietze. 2014. A Taxonomy of Microtasks on the Web. In: *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, 218–223. HT '14. New York, NY, USA: ACM. doi:10.1145/2631775.2631819, <http://doi.acm.org/10.1145/2631775.2631819> (zugegriffen: 15. Juni 2019).
- Gadiraju, Ujwal, Ricardo Kawase, Stefan Dietze und Gianluca Demartini. 2015. Understanding Malicious Behavior in Crowdsourcing Platforms: The Case of Online Surveys. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1631–1640. CHI '15. New York, NY, USA: ACM. doi:10.1145/2702123.2702443, <http://doi.acm.org/10.1145/2702123.2702443> (zugegriffen: 6. November 2015).

- Gambarotto, Francesca und Alberto Cammozzo. 2010. Dreams of silence: Employee voice and innovation in a public sector community of practice. *Innovation* 12, Nr. 2 (1. August): 166–179. doi:10.5172/impp.12.2.166.
- Garrison, D. R. 1991. Critical thinking and adult education: a conceptual model for developing critical thinking in adult learners. *International Journal of Lifelong Education* 10, Nr. 4 (1. Oktober): 287–303. doi:10.1080/0260137910100403.
- Ge, Xun und Susan M. Land. 2004. A conceptual framework for scaffolding III-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development* 52, Nr. 2 (1. Juni): 5–22. doi:10.1007/BF02504836.
- Gerlach, Sophia und Alexander Brem. 2017. Idea management revisited: A review of the literature and guide for implementation. *International Journal of Innovation Studies* 1, Nr. 2 (1. November): 144–161. doi:10.1016/j.ijis.2017.10.004.
- Germine, Laura, Ken Nakayama, Bradley C. Duchaine, Christopher F. Chabris, Garga Chatterjee und Jeremy B. Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review* 19, Nr. 5 (1. Oktober): 847–857. doi:10.3758/s13423-012-0296-9.
- Ghaye, Tony. 2010. A Reflective Inquiry as Participatory and Appreciative Action and Reflection. In: *Handbook of Reflection and Reflective Inquiry: Mapping a Way of Knowing for Professional Reflective Inquiry*, hg. von Nona Lyons, 553–569. Boston, MA: Springer US. doi:10.1007/978-0-387-85744-2\_28, [https://doi.org/10.1007/978-0-387-85744-2\\_28](https://doi.org/10.1007/978-0-387-85744-2_28) (zugegriffen: 16. Juni 2019).
- Glazer, Courtney, Lynda Abbott und Judith Harris. 2004. A teacher-developed process for collaborative professional reflection. *Reflective Practice* 5, Nr. 1 (1. Februar): 33–46. doi:10.1080/1462394032000169947.
- Glogger, Inga, Lars Holzäpfel, Rolf Schwonke, Matthias Nückles und Alexander Renkl. 2009. Activation of Learning Strategies in Writing Learning Journals: The Specificity of Prompts Matters. *Zeitschrift für Pädagogische Psychologie* 23, Nr. 2 (1. Januar): 95–104. doi:10.1024/1010-0652.23.2.95.
- Gosling, Samuel D., Simine Vazire, Sanjay Srivastava und Oliver P. John. 2004. Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist* 59, Nr. 2: 93–104. doi:10.1037/0003-066X.59.2.93.
- Gray, Bette. 2004. Informal Learning in an Online Community of Practice. *Journal of Distance Education* 19, Nr. 1: 20–35.
- Griffith, Brian A. und Gina Frieden. 2000. Facilitating Reflective Thinking in Counselor Education. *Counselor Education and Supervision* 40, Nr. 2: 82–93. doi:10.1002/j.1556-6978.2000.tb01240.x.
- Gustafsson, Christine und Ingegerd Fagerberg. 2004. Reflection, the way to professional development? *Journal of Clinical Nursing* 13, Nr. 3 (1. März): 271–280. doi:10.1046/j.1365-2702.2003.00880.x.
- Gutwin, Carl, Andy Cockburn und Ashley Coveney. 2017. Peripheral Popout: The Influence of Visual Angle and Stimulus Intensity on Popout Effects. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 208–219. CHI '17. New York, NY, USA: ACM. doi:10.1145/3025453.3025984, <http://doi.acm.org/10.1145/3025453.3025984>.
- Hamann, Janet M. 2002. *Reflective Practices and Confluent Educational Perspectives: Three Exploratory Studies*. <https://eric.ed.gov/?id=ED472393> (zugegriffen: 7. Juli 2019).

- Hara, Noriko, Curtis Jay Bonk und Charoula Angeli. 2000. Content analysis of online discussion in an applied educational psychology course. *Instructional Science* 28, Nr. 2 (1. März): 115–152. doi:10.1023/A:1003764722829.
- Hartley, James und Lucy R. Betts. 2010. Four layouts and a finding: the effects of changes in the order of the verbal labels and numerical values on Likert-type scales. *International Journal of Social Research Methodology* 13, Nr. 1 (1. Februar): 17–27. doi:10.1080/13645570802648077.
- Hatton, Neville und David Smith. 1995. Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education* 11, Nr. 1 (Januar): 33–49. doi:10.1016/0742-051X(94)00012-U.
- Hauser, David J. und Norbert Schwarz. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods* 48, Nr. 1 (1. März): 400–407. doi:10.3758/s13428-015-0578-z.
- Hawkey, Kate. 1995. Learning from Peers: The Experience of Student Teachers in School-Based Teacher Education. *Journal of Teacher Education* 46, Nr. 3: 175–183.
- Hayes, Andrew F. und Klaus Krippendorff. 2007. Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures* 1, Nr. 1 (1. April): 77–89. doi:10.1080/19312450709336664.
- Heer, Jeffrey und Michael Bostock. 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 203–212. CHI '10. New York, NY, USA: ACM. doi:10.1145/1753326.1753357, <http://doi.acm.org/10.1145/1753326.1753357> (zugegriffen: 5. November 2015).
- Heilman, Michael und Noah A. Smith. 2010. Rating Computer-generated Questions with Mechanical Turk. In: *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, 35–40. CSLDAMT '10. Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1866696.1866701> (zugegriffen: 27. Mai 2019).
- Herrmann, Thomas und Jan Nierhoff. 2017. Prompting – A Feature of General Relevance in HCI-Supported Task Workflows. In: *HCI International 2017 – Posters' Extended Abstracts*, 123–129. Communications in Computer and Information Science. Springer, Cham, 9. Juli. doi:10.1007/978-3-319-58750-9\_17, [https://link.springer.com/chapter/10.1007/978-3-319-58750-9\\_17](https://link.springer.com/chapter/10.1007/978-3-319-58750-9_17) (zugegriffen: 3. Januar 2018).
- Hettich, Paul. 1990. Journal Writing: Old Fare or Nouvelle Cuisine? *Teaching of Psychology* 17, Nr. 1 (1. Februar): 36–39. doi:10.1207/s15328023top1701\_8.
- Hoegl, Martin und K. Praveen Parboteeah. 2006. Team reflexivity in innovative projects. *R&D Management* 36, Nr. 2: 113–125. doi:10.1111/j.1467-9310.2006.00420.x.
- Hoffman, Bobby und Alexandru Spataru. 2008. The influence of self-efficacy and metacognitive prompting on math problem-solving efficiency. *Contemporary Educational Psychology* 33, Nr. 4 (Oktober): 875–893. doi:10.1016/j.cedpsych.2007.07.002.
- Holliday, William G. 1983. Overprompting science students using adjunct study questions. *Journal of Research in Science Teaching* 20, Nr. 3: 195–201. doi:10.1002/tea.3660200303.



- Horton, John J., David G. Rand und Richard J. Zeckhauser. 2011. The online laboratory: conducting experiments in a real labor market. *Experimental Economics* 14, Nr. 3 (1. September): 399–425. doi:10.1007/s10683-011-9273-9.
- Howe, Jeff. 2006. The Rise of Crowdsourcing. *Wired*, 1. Juni. <https://www.wired.com/wired/archive/14.06/crowds.html> (zugegriffen: 24. Mai 2019).
- . 2016. Crowdsourcing: A Definition. *Crowdsourcing*. 6. Februar. [https://www.crowdsourcing.com/cs/2006/06/crowdsourcing\\_a.html](https://www.crowdsourcing.com/cs/2006/06/crowdsourcing_a.html) (zugegriffen: 24. Mai 2019).
- Høyrup, Steen. 2004. Reflection as a core process in organisational learning. *Journal of Workplace Learning* 16, Nr. 8: 442–454.
- Ifenthaler, Dirk. 2012. Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society* 15, Nr. 1 (Januar): 38–52.
- Introne, Joshua, Bryan Semaan und Sean Goggins. 2016. A Sociotechnical Mechanism for Online Support Provision. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3559–3571. CHI '16. New York, NY, USA: ACM. doi:10.1145/2858036.2858582, <http://doi.acm.org/10.1145/2858036.2858582> (zugegriffen: 17. Juni 2016).
- Ipeirotis, Panagiotis G. 2010. Demographics of Mechanical Turk. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. <http://papers.ssrn.com/abstract=1585030> (zugegriffen: 18. November 2015).
- Isaacs, Ellen, Artie Konrad, Alan Walendowski, Thomas Lennig, Victoria Hollis und Steve Whittaker. 2013. Echoes from the Past: How Technology Mediated Reflection Improves Well-being. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1071–1080. CHI '13. New York, NY, USA: ACM. doi:10.1145/2470654.2466137, <http://doi.acm.org/10.1145/2470654.2466137> (zugegriffen: 26. Oktober 2015).
- Jackson, Corey Brian, Kevin Crowston, Gabriel Mugar und Carsten Østerlund. 2016. „Guess What! You'Re the First to See This Event“: Increasing Contribution to Online Production Communities. In: *Proceedings of the 19th International Conference on Supporting Group Work*, 171–179. GROUP '16. New York, NY, USA: ACM. doi:10.1145/2957276.2957284, (zugegriffen: 29. November 2016).
- Jae Jeong, Heon. 2016. The level of collapse we are allowed: Comparison of different response scales in Safety Attitudes Questionnaire. *Biometrics & Biostatistics International Journal* 4, Nr. 4 (12. September). doi:10.15406/bbij.2016.04.00100, <http://medcraveonline.com/BBIJ/BBIJ-04-00100.php> (zugegriffen: 17. Oktober 2019).
- Jamieson, Susan. 2004. Likert scales: how to (ab)use them. *Medical Education* 38, Nr. 12: 1217–1218. doi:10.1111/j.1365-2929.2004.02012.x.
- Järvinen, Annikki und Esa Poikela. 2001. Modelling reflective and contextual learning at work. *Journal of Workplace Learning* 13, Nr. 7/8 (1. Dezember): 282–290. doi:10.1108/13665620110411067.
- Johnson, Christopher M. 2001. A survey of current research on online communities of practice. *The Internet and Higher Education* 4, Nr. 1: 45–60. doi:10.1016/S1096-7516(01)00047-1.
- Johnson, Eric J., Suzanne B. Shu, Benedict G. C. Dellaert, Craig Fox, Daniel G. Goldstein, Gerald Häubl, Richard P. Larrick, u. a. 2012. Beyond nudges: Tools of a choice architecture. *Marketing Letters* 23, Nr. 2 (1. Juni): 487–504. doi:10.1007/s11002-012-9186-1.

- Johnston, Andrew, Shigeki Amitani und Ernest Edmonds. 2005. Amplifying reflective thinking in musical performance. In: *Proceedings of the 5th conference on Creativity & cognition*, 166–175. ACM.
- Johri, Aditya und Seungwon Yang. 2017. Scaffolded Help for Learning: How Experts Collaboratively Support Newcomer Participation in Online Communities. In: *Proceedings of the 8th International Conference on Communities and Technologies*, 149–158. C&T '17. New York, NY, USA: ACM. doi:10.1145/3083671.3083694, <http://doi.acm.org/10.1145/3083671.3083694>.
- Jones, M. Cameron und Elizabeth F. Churchill. 2009. Conversations in Developer Communities: A Preliminary Analysis of the Yahoo! Pipes Community. In: *Proceedings of the Fourth International Conference on Communities and Technologies*, 195–204. C&T '09. New York, NY, USA: ACM. doi:10.1145/1556460.1556489, <http://doi.acm.org/10.1145/1556460.1556489> (zugegriffen: 3. Juni 2016).
- Kaczynski, Andrew T., Sonja A. Wilhelm Stanis und J. Aaron Hipp. 2014. Point-of-decision prompts for increasing park-based physical activity: A crowdsourcing analysis. *Preventive Medicine* 69 (Dezember): 87–89. doi:10.1016/j.ypmed.2014.08.029.
- Kalyuga, Slava. 2009. The Expertise Reversal Effect. *Managing Cognitive Load in Adaptive Multimedia Learning*: 58–80. doi:10.4018/978-1-60566-048-6.ch003.
- Karmiloff-Smith, Annette und Bärbel Inhelder. 1974. If you want to get ahead, get a theory. *Cognition* 3, Nr. 3 (1. Januar): 195–212. doi:10.1016/0010-0277(74)90008-0.
- Kazai, Gabriella, Jaap Kamps und Natasa Milic-Frayling. 2012. The Face of Quality in Crowdsourcing Relevance Labels: Demographics, Personality and Labeling Accuracy. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 2583–2586. CIKM '12. New York, NY, USA: ACM. doi:10.1145/2396761.2398697, (zugegriffen: 12. November 2015).
- Keith, Nina und Michael Frese. 2005. Self-Regulation in Error Management Training: Emotion Control and Metacognition as Mediators of Performance Effects. *Journal of Applied Psychology* 90, Nr. 4: 677–691. doi:10.1037/0021-9010.90.4.677.
- Kellas, Jody Koenig und April R. Trees. 2006. Finding Meaning in Difficult Family Experiences: Sense-Making and Interaction Processes During Joint Family Storytelling. *Journal of Family Communication* 6, Nr. 1 (1. Januar): 49–76. doi:10.1207/s15327698jfc0601\_4.
- Kemmis, Stephen. 1985. Action research and the politics of reflection. In: *Reflection: Turning experience into learning*, hg. von David Boud, Rosemary Keogh, und David Walker, 139–163.
- Kim, Dongsik und Seunghee Lee. 2002. Designing Collaborative Reflection Supporting Tools in e-Project-Based Learning Environments. *Journal of Interactive Learning Research* 13, Nr. 4: 375–392.
- Kim, Hesook Suzie, Laurie M. Lauzon Clabo, Patricia Burbank, Mary Leveillee und Diane Martins. 2010. Application of Critical Reflective Inquiry in Nursing Education. In: *Handbook of Reflection and Reflective Inquiry: Mapping a Way of Knowing for Professional Reflective Inquiry*, hg. von Nona Lyons, 159–172. Boston, MA: Springer US. doi:10.1007/978-0-387-85744-2\_8, [https://doi.org/10.1007/978-0-387-85744-2\\_8](https://doi.org/10.1007/978-0-387-85744-2_8) (zugegriffen: 16. Juni 2019).

- Kimmerle, Joachim, Martina Bientzle und Ulrike Cress. 2014. Personal Experiences and Emotionality in Health-Related Knowledge Exchange in Internet Forums: A Randomized Controlled Field Experiment Comparing Responses to Facts Vs Personal Experiences. *Journal of Medical Internet Research* 16, Nr. 12 (4. Dezember): e277. doi:10.2196/jmir.3766.
- Kimmerle, Joachim und Ulrike Cress. 2009. Knowledge Communication with Shared Databases. *Database Technologies: Concepts, Methodologies, Tools, and Applications*: 1670–1681. doi:10.4018/978-1-60566-058-5.ch102.
- Kimmerle, Joachim, Katrin Wodzicki, Halszka Jarodzka und Ulrike Cress. 2011. Value of Information, Behavioral Guidelines, and Social Value Orientation in an Information-Exchange Dilemma. *Group Dynamics: Theory, Research, and Practice* 15, Nr. 2 (1. Juni): 173–186. doi:10.1037/a0021467.
- King, Alison. 1994. Guiding Knowledge Construction in the Classroom: Effects of Teaching Children How to Question and How to Explain. *American Educational Research Journal* 31, Nr. 2 (20. Juni): 338–368. doi:10.3102/00028312031002338.
- King, Alison und Barak Rosenshine. 1993. Effects of Guided Cooperative Questioning on Children's Knowledge Construction. *The Journal of Experimental Education* 61, Nr. 2 (1. Januar): 127–148. doi:10.1080/00220973.1993.9943857.
- Kittur, Aniket, Ed H. Chi und Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 453–456. CHI '08. New York, NY, USA: ACM. doi:10.1145/1357054.1357127, <http://doi.acm.org/10.1145/1357054.1357127> (zugegriffen: 8. November 2015).
- Kleemann, Frank, G. Günter Voß und Kerstin Rieder. 2008. Un(der)paid Innovators: The Commercial Utilization of Consumer Work through Crowdsourcing. *Science, Technology & Innovation Studies* 4, Nr. 1: 22.
- Knipfer, Kristin, Barbara Kump, Daniel Wessel und Ulrike Cress. 2013. Reflection as a catalyst for organisational learning. *Studies in Continuing Education* 35, Nr. 1 (1. März): 30–48. doi:10.1080/0158037X.2012.683780.
- Kocielnik, Rafal, Lillian Xiao, Daniel Avrahami und Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, Nr. 2 (Juli): 70:1–70:26. doi:10.1145/3214273.
- Kolb, Alice Y. und David A. Kolb. 2005. Learning Styles and Learning Spaces: Enhancing Experiential Learning in Higher Education. *Academy of Management Learning & Education* 4, Nr. 2 (1. Juni): 193–212. doi:10.5465/amle.2005.17268566.
- . 2018. Eight important things to know about the experiential learning cycle. *Australian Educational Leader* 40, Nr. 3 (September): 8.
- Kolb, David A. 1984. *Experiential Learning: Experience as the Source of Learning and Development*. Prentice-Hall.
- . 2007. *The Kolb learning style inventory*. Hay Resources Direct Boston, MA.
- Kolb, David A. und Ronald Fry. 1975. Towards an Applied Theory of Experiential Learning. In: *Theories of group processes*, hg. von Cary Lynn Cooper. London ; New York: Wiley & Sons. <https://trove.nla.gov.au/version/45413677> (zugegriffen: 7. Juni 2019).
- Komarov, Steven, Katharina Reinecke und Krzysztof Z. Gajos. 2013. Crowdsourcing Performance Evaluations of User Interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 207–216. CHI '13. New

- York, NY, USA: ACM. doi:10.1145/2470654.2470684, <http://doi.acm.org/10.1145/2470654.2470684> (zugegriffen: 6. November 2015).
- Kori, Külli, Margus Pedaste, Äli Leijen und Mario Mäeots. 2014. Supporting reflection in technology-enhanced learning. *Educational Research Review* 11 (Januar): 45–55. doi:10.1016/j.edurev.2013.11.003.
- Krippendorff, Klaus. 2004. Reliability in Content Analysis.: Some Common Misconceptions and Recommendations. *Human Communication Research* 30, Nr. 3 (Juli): 411–433. doi:10.1111/j.1468-2958.2004.tb00738.x.
- . 2011. Agreement and information in the reliability of coding. *Communication Methods and Measures* 5, Nr. 2: 93–112.
- Krogstie, Birgit R., Michael Prilla und Viktoria Pammer. 2013. Understanding and Supporting Reflective Learning Processes in the Workplace: The CSRL Model. In: *Scaling up Learning for Sustained Impact*, hg. von Davinia Hernández-Leo, Tobias Ley, Ralf Klamma, und Andreas Harrer, 151–164. Lecture Notes in Computer Science 8095. Springer Berlin Heidelberg, 1. Januar. [http://link.springer.com/chapter/10.1007/978-3-642-40814-4\\_13](http://link.springer.com/chapter/10.1007/978-3-642-40814-4_13) (zugegriffen: 18. September 2014).
- Krosnick, Jon A. und Leandre R. Fabrigar. 2012. Designing Rating Scales for Effective Measurement in Surveys. In: *Survey Measurement and Process Quality*, 141–164. John Wiley & Sons, Ltd. doi:10.1002/9781118490013.ch6, <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118490013.ch6> (zugegriffen: 15. Juni 2019).
- Kudrowitz, Barry Matthew und David Wallace. 2013. Assessing the quality of ideas from prolific, early-stage product ideation. *Journal of Engineering Design* 24, Nr. 2 (1. Februar): 120–139. doi:10.1080/09544828.2012.676633.
- Lai, Guolin und Brendan Calandra. 2007. Using Online Scaffolds to Enhance Preservice Teachers' Reflective Journal Writing: A Qualitative Analysis. *International Journal of Technology in Teaching and Learning* 3, Nr. 3: 66–81.
- Lakhani, Karim R, Lars Bo Jeppesen, Peter A Lohse und Jill A Panetta. 2007. The Value of Openness in Scientific Problem Solving: 58.
- Land, Susan M. und Carla Zembal-Saul. 2003. Scaffolding reflection and articulation of scientific explanations in a data-rich, project-based learning environment: An investigation of progress portfolio. *Educational Technology Research and Development* 51, Nr. 4 (Dezember): 65–84. doi:10.1007/BF02504544.
- Lave, Jean und Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Lee, Seung-hee. 2005. Design and analysis of reflection-supporting tools in computer supported collaborative learning. *International Journal of Instructional Technology and Distance Learning* 2, Nr. 3: 49–56.
- Leijen, Äli, Ineke Lam, Liesbeth Wildschut, P. Robert-Jan Simons und Wilfried Admiraal. 2009. Streaming video to enhance students' reflection in dance education. *Computers & Education* 52, Nr. 1 (1. Januar): 169–176. doi:10.1016/j.compedu.2008.07.010.
- Lesser, E. L. und J. Storck. 2001. Communities of practice and organizational performance. *IBM Systems Journal* 40, Nr. 4: 831–841. doi:10.1147/sj.404.0831.
- Levina, Natalia. 2005. Collaborating on Multiparty Information Systems Development Projects: A Collective Reflection-in-Action View. *Information Systems Research* 16, Nr. 2 (1. Juni): 109–130. doi:10.1287/isre.1050.0055.

- Lewin, Kurt. 1951. *Field theory in social science: selected theoretical papers* (Edited by Dorwin Cartwright.). Field theory in social science: selected theoretical papers (Edited by Dorwin Cartwright.). Oxford, England: Harpers.
- Lin, Xiaodong, Cindy Hmelo, Charles K. Kinzer und Teresa J. Secules. 1999. Designing technology to support reflection. *Educational Technology Research and Development* 47, Nr. 3 (1. September): 43–62. doi:10.1007/BF02299633.
- Lin, Xiaodong und James D. Lehman. 1999. Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching* 36, Nr. 7: 837–858. doi:10.1002/(SICI)1098-2736(199909)36:7<837::AID-TEA6>3.0.CO;2-U.
- Lindgren, Ida und Gabriella Jansson. 2013. Electronic services in the public sector: A conceptual framework. *Government Information Quarterly* 30, Nr. 2 (April): 163–172. doi:10.1016/j.giq.2012.10.005.
- Lisewski, Bernard. 2004. Implementing a Learning Technology Strategy: Top-Down Strategy Meets Bottom-Up Culture. *ALT-J: Research in Learning Technology* 12, Nr. 2 (Juni): 175–188. doi:10.1080/0968776042000216228.
- Loo, Robert und Karran Thorpe. 2002. Using reflective learning journals to improve individual and team performance. *Team Performance Management: An International Journal* 8, Nr. 5/6 (1. September): 134–139. doi:10.1108/13527590210442258.
- Lupton, Deborah. 2016. *The Quantified Self*. John Wiley & Sons.
- MacDonald, Ronald J. 2008. Professional Development for Information Communication Technology Integration. *Journal of Research on Technology in Education* 40, Nr. 4 (1. Juni): 429–445. doi:10.1080/15391523.2008.10782515.
- Magnusson, Peter R. 2009. Exploring the Contributions of Involving Ordinary Users in Ideation of Technology-Based Services. *Journal of Product Innovation Management* 26, Nr. 5 (September): 578–593. doi:10.1111/j.1540-5885.2009.00684.x.
- Magnusson, Peter R., Johan Netz und Erik Wästlund. 2014. Exploring holistic intuitive idea screening in the light of formal criteria. *Technovation* 34, Nr. 5. Special Issue: Base of the Pyramid (1. Mai): 315–326. doi:10.1016/j.technovation.2014.03.003.
- Mamykina, Lena, Elizabeth M. Heitkemper, Arlene M. Smaldone, Rita Kukafka, Heather Cole-Lewis, Patricia G. Davidson, Elizabeth D. Mynatt, u. a. 2016. Structured scaffolding for reflection and problem solving in diabetes self-management: qualitative study of mobile diabetes detective. *Journal of the American Medical Informatics Association* 23, Nr. 1 (1. Januar): 129–136. doi:10.1093/jamia/ocv169.
- Mamykina, Lena, Elizabeth Mynatt, Patricia Davidson und Daniel Greenblatt. 2008. MAHI: Investigation of Social Scaffolding for Reflective Thinking in Diabetes Management. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 477–486. CHI '08. New York, NY, USA: ACM. doi:10.1145/1357054.1357131, <http://doi.acm.org/10.1145/1357054.1357131> (zugegriffen: 11. Februar 2015).
- Manouselis, Nikos, Hendrik Drachsler, Katrien Verbert und Erik Duval. 2012. *Recommender Systems for Learning*. Springer Science & Business Media.
- Marcu, Gabriela, Anind K Dey und Sara Kiesler. 2014. Designing for Collaborative Reflection. In: . Oldenburg, Germany.
- Martin, David, Benjamin V. Hanrahan, Jacki O'Neill und Neha Gupta. 2014. Being a Turker. In: *Proceedings of the 17th ACM Conference on Computer Supported*

- Cooperative Work & Social Computing*, 224–235. CSCW '14. New York, NY, USA: ACM. doi:10.1145/2531602.2531663, <http://doi.acm.org/10.1145/2531602.2531663> (zugegriffen: 9. November 2015).
- Mason, Winter und Siddharth Suri. 2011. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods* 44, Nr. 1 (30. Juni): 1–23. doi:10.3758/s13428-011-0124-6.
- Mason, Winter und Duncan J. Watts. 2009. Financial Incentives and the „Performance of Crowds“. In: *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 77–85. HCOMP '09. New York, NY, USA: ACM. doi:10.1145/1600150.1600175, <http://doi.acm.org/10.1145/1600150.1600175> (zugegriffen: 27. Mai 2019).
- McHugh, Mary L. 2013. The Chi-square test of independence. *Biochemia Medica* 23, Nr. 2 (15. Juni): 143–149. doi:10.11613/BM.2013.018.
- McKenzie-Mohr, Doug. 2000. Fostering sustainable behavior through community-based social marketing. *American Psychologist* 55, Nr. 5: 531–537. doi:10.1037/0003-066X.55.5.531.
- McLure Wasko, M. und S. Faraj. 2000. “It is what one does”: why people participate and help others in electronic communities of practice. *The Journal of Strategic Information Systems* 9, Nr. 2 (1. September): 155–173. doi:10.1016/S0963-8687(00)00045-7.
- McNicol, Sarah, Cathy Lewin, Anna Keune und Tarmo Toikkanen. 2014. Facilitating Student Reflection through Digital Technologies in the iTEC Project: Pedagogically-Led Change in the Classroom. In: *Learning and Collaboration Technologies. Technology-Rich Environments for Learning and Collaboration*, hg. von Panayiotis Zaphiris und Andri Ioannou, 297–308. Lecture Notes in Computer Science 8524. Springer International Publishing, 22. Juni. doi:10.1007/978-3-319-07485-6\_30, [http://link.springer.com/chapter/10.1007/978-3-319-07485-6\\_30](http://link.springer.com/chapter/10.1007/978-3-319-07485-6_30) (zugegriffen: 20. November 2015).
- Mekler, Elisa D., Ioanna Iacovides und Julia Ayumi Bopp. 2018. „A Game That Makes You Question...“: Exploring the Role of Reflection for the Player Experience. In: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 315–327. CHI PLAY '18. New York, NY, USA: ACM. doi:10.1145/3242671.3242691, <http://doi.acm.org/10.1145/3242671.3242691> (zugegriffen: 29. Mai 2019).
- Mezirow, Jack. 1990. How Critical Reflection Triggers Transformative Learning. In: *Fostering critical reflection in adulthood: a guide to transformative and emancipatory learning*, 1–20. Jossey-Bass Publishers.
- Mirsch, Tobias, Christiane Lehrer und Reinhard Jung. 2017. Digital Nudging: Altering User Behavior in Digital Environments. In: *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*, 634–648. St. Gallen, Switzerland, 14. Februar. <https://wi2017.ch/de/proceedings> (zugegriffen: 29. Juni 2019).
- Molenaar, Inge, Peter Sleegers und Carla van Boxtel. 2014. Metacognitive scaffolding during collaborative learning: a promising combination. *Metacognition and Learning* 9, Nr. 3 (17. Juli): 309–332. doi:10.1007/s11409-014-9118-y.
- Mols, Ine, Elise van den Hoven und Berry Eggen. 2016a. Technologies for Everyday Life Reflection: Illustrating a Design Space. In: *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, 53–61. TEI '16. New York, NY, USA: ACM. doi:10.1145/2839462.2839466, <http://doi.acm.org/10.1145/2839462.2839466> (zugegriffen: 29. Mai 2019).

- . 2016b. Informing Design for Reflection: An Overview of Current Everyday Practices. In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, 21:1–21:10. NordiCHI '16. New York, NY, USA: ACM. doi:10.1145/2971485.2971494, <http://doi.acm.org/10.1145/2971485.2971494> (zugegriffen: 2. Dezember 2016).
- Montero, Calkin A. S. und Kenji Araki. 2005. Enhancing Computer Chat: Toward a Smooth User-computer Interaction. In: *Proceedings of the 9th International Conference on Knowledge-Based Intelligent Information and Engineering Systems - Volume Part I*, 918–924. KES'05. Berlin, Heidelberg: Springer-Verlag. doi:10.1007/11552413\_131, [http://dx.doi.org/10.1007/11552413\\_131](http://dx.doi.org/10.1007/11552413_131) (zugegriffen: 12. August 2019).
- Moon, Jae Yun und Lee S. Sproull. 2008. The Role of Feedback in Managing the Internet-Based Volunteer Work Force. *Information Systems Research* 19, Nr. 4 (1. Dezember): 494–515. doi:10.1287/isre.1080.0208.
- Moon, Jennifer A. 1999. *Reflection in learning and professional development: Theory and practice*. Psychology Press.
- . 2006a. *A handbook of reflective and experiential learning: theory and practice*. London; New York: Routledge. <http://www.myilibrary.com?id=5227> (zugegriffen: 5. Juni 2019).
- . 2006b. *Learning Journals: A Handbook for Reflective Practice and Professional Development*. Routledge.
- Morris, Jane und Graham Stew. 2007. Collaborative reflection: how far do 2:1 models of learning in the practice setting promote peer reflection? *Reflective Practice* 8, Nr. 3 (1. August): 419–432. doi:10.1080/14623940701425220.
- Murphy, Marian, Maria Dempsey und Carmel Halton. 2010. Reflective Inquiry in Social Work Education. In: *Handbook of Reflection and Reflective Inquiry: Mapping a Way of Knowing for Professional Reflective Inquiry*, hg. von Nona Lyons, 173–188. Boston, MA: Springer US. doi:10.1007/978-0-387-85744-2\_9, [https://doi.org/10.1007/978-0-387-85744-2\\_9](https://doi.org/10.1007/978-0-387-85744-2_9) (zugegriffen: 16. Juni 2019).
- Neill, Sarah J. 2006. Grounded theory sampling: The contribution of reflexivity. *Journal of Research in Nursing* 11, Nr. 3 (1. Mai): 253–260. doi:10.1177/1744987106051850.
- Nguyen, Quoc Dinh, Nicolas Fernandez, Thierry Karsenti und Bernard Charlin. 2014. What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Medical Education* 48, Nr. 12 (1. Dezember): 1176–1189. doi:10.1111/medu.12583.
- Nicholson, Sheila A., Nathan Bond, Sheila A. Nicholson und Nathan Bond. 2003. Collaborative Reflection and Professional Community Building: An Analysis of Preservice Teachers' Use of an Electronic Discussion Board. *Journal of Technology and Teacher Education* 11, Nr. 2: 259–279.
- Nissilä, Sade-Pirkko. 2005. Individual and collective reflection: how to meet the needs of development in teaching. *European Journal of Teacher Education* 28, Nr. 2 (1. Juni): 209–219. doi:10.1080/02619760500093354.
- Nocon, Marc, Falk Müller-Riemenschneider, Katleen Nitzschke und Stefan N. Willich. 2010. Increasing physical activity with point-of-choice prompts – a systematic review. *Scandinavian Journal of Public Health* (2. Juli). doi:10.1177/1403494810375865, <http://sjp.sagepub.com/content/early/2010/07/01/1403494810375865> (zugegriffen: 17. Februar 2016).

- Nowak, Stefanie und Stefan Rüger. 2010. How Reliable Are Annotations via Crowdsourcing: A Study About Inter-annotator Agreement for Multi-label Image Annotation. In: *Proceedings of the International Conference on Multimedia Information Retrieval*, 557–566. MIR '10. New York, NY, USA: ACM. doi:10.1145/1743384.1743478, <http://doi.acm.org/10.1145/1743384.1743478> (zugegriffen: 21. Mai 2019).
- Nückles, Matthias, Sandra Hübner und Alexander Renkl. 2006. *The Pitfalls of Overprompting in Writing-to-Learn*.
- van Oldenbeek, Maxim, Till Winkler, Julie Buhl-Wiggers und Daniel Hardt. 2019. Nudging in Blended Learning: Evaluation of Email-Based Progress Feedback in a Flipped-Classroom Information Systems Course. *Research Papers* (15. Mai). [https://aisel.aisnet.org/ecis2019\\_rp/186](https://aisel.aisnet.org/ecis2019_rp/186).
- Oleson, David, Alexander Sorokin, Greg Laughlin, Vaughn Hester, John Le und Lukas Biewald. 2011. Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing. In: *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*. 24. August. <https://www.aaai.org/ocs/index.php/WS/AAAIW11/paper/view/3995> (zugegriffen: 27. Mai 2019).
- Oliveira, I., L. Tinoca und A. Pereira. 2011. Online group work patterns: How to promote a successful collaboration. *Computers & Education* 57, Nr. 1 (1. August): 1348–1357. doi:10.1016/j.compedu.2011.01.017.
- Oliver, R., A. Omari und J. Herrington. 1998. Exploring student interactions in collaborative World Wide Web learning environments. *Journal of Educational Multimedia and Hypermedia* 7: 263–287.
- Oppenheimer, Daniel M., Tom Meyvis und Nicolas Davidenko. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology* 45, Nr. 4 (1. Juli): 867–872. doi:10.1016/j.jesp.2009.03.009.
- Orlikowski, Wanda J. 2008. Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. In: *Resources, Co-Evolution and Artifacts*, 255–305. Computer Supported Cooperative Work. Springer London. doi:10.1007/978-1-84628-901-9\_10, [http://link.springer.com/chapter/10.1007/978-1-84628-901-9\\_10](http://link.springer.com/chapter/10.1007/978-1-84628-901-9_10) (zugegriffen: 30. Januar 2017).
- Osborn, Alex F. 1953. *Applied imagination: Principles and procedures of creative problem-solving*. C. Scribner's Sons. <http://www.amazon.com/Applied-imagination-Principles-procedures-problem-solving/dp/B0007HKJUM>.
- Palanque, Philippe, Marco Winckler und Célia Martinie. 2011. A Formal Model-Based Approach for Designing Interruptions-Tolerant Advanced User Interfaces. In: *Model-Driven Development of Advanced User Interfaces*, hg. von Heinrich Hussmann, Gerrit Meixner, und Detlef Zuehlke, 143–169. Studies in Computational Intelligence 340. Springer Berlin Heidelberg, 1. Januar. [http://link.springer.com/chapter/10.1007/978-3-642-14562-9\\_8](http://link.springer.com/chapter/10.1007/978-3-642-14562-9_8) (zugegriffen: 3. November 2014).
- Pallant, Julie F. 2011. *SPSS Survival Manual: A step by step guide to data analysis using SPSS*. Crows Nest, NSW: Allen & Unwin.
- Paolacci, Gabriele und Jesse Chandler. 2014. Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science* 23, Nr. 3 (1. Juni): 184–188. doi:10.1177/0963721414531598.



- Paolacci, Gabriele, Jesse Chandler und Panagiotis G. Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, Nr. 5: 9.
- Papadopoulos, Pantelis M., Stavros N. Demetriadis, Ioannis G. Stamelos und Ioannis A. Tsoukalas. 2009. Prompting students' context-generating cognitive activity in ill-structured domains: does the prompting mode affect learning? *Educational Technology Research and Development* 57, Nr. 2 (1. April): 193–210. doi:10.1007/s11423-008-9105-6.
- Pea, Roy D. 2004. The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *Journal of the Learning Sciences* 13, Nr. 3 (1. Juli): 423–451. doi:10.1207/s15327809jls1303\_6.
- Pedersen, J., D. Kocsis, A. Tripathi, A. Tarrell, A. Weerakoon, N. Tahmasbi, J. Xiong, W. Deng, O. Oh und G. de Vreede. 2013. Conceptual Foundations of Crowdsourcing: A Review of IS Research. In: *2013 46th Hawaii International Conference on System Sciences*, 579–588. Januar. doi:10.1109/HICSS.2013.143.
- Pennebaker, J.W., R.L. Boyd, K. Jordan und K. Blackburn. 2015. The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin. DOI: 10.15781/T29G6Z.
- Peters, Vanessa L. und Jim Hewitt. 2010. An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education* 54, Nr. 4 (1. Mai): 951–961. doi:10.1016/j.compedu.2009.09.030.
- Piaget, Jean. 1976. Piaget's Theory. In: *Piaget and His School: A Reader in Developmental Psychology*, hg. von Bärbel Inhelder, Harold H. Chipman, und Charles Zwingmann, 11–23. Springer Study Edition. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-46323-5\_2, [https://doi.org/10.1007/978-3-642-46323-5\\_2](https://doi.org/10.1007/978-3-642-46323-5_2) (zugegriffen: 19. Oktober 2019).
- Pirzadeh, Afarin, Li He und Erik Stolterman. 2013. Personal Informatics and Reflection: A Critical Examination of the Nature of Reflection. In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 1979–1988. CHI EA '13. New York, NY, USA: ACM. doi:10.1145/2468356.2468715, <http://doi.acm.org/10.1145/2468356.2468715> (zugegriffen: 25. Juni 2015).
- Platzer, Hazel, Jannice Snelling und David Blake. 1997. Promoting Reflective Practitioners in Nursing: a review of theoretical models and research into the use of diaries and journals to facilitate reflection. *Teaching in Higher Education* 2, Nr. 2 (1. Juni): 103–121. doi:10.1080/1356251970020202.
- Ploderer, Bernd, Wolfgang Reitberger, Harri Oinas-Kukkonen und Julia van Gemert-Pijnen. 2014. Social interaction and reflection for behaviour change. *Personal and Ubiquitous Computing* 18, Nr. 7 (4. Juli): 1667–1676. doi:10.1007/s00779-014-0779-y.
- Poetz, Marion K. und Martin Schreier. 2012. The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?: The Value of Crowdsourcing. *Journal of Product Innovation Management* 29, Nr. 2 (März): 245–256. doi:10.1111/j.1540-5885.2011.00893.x.
- Polanyi, Michael. 1966. *The Tacit Dimension*. University of Chicago Press.
- Poldner, Eric, Marieke Van der Schaaf, P. Robert-Jan Simons, Jan Van Tartwijk und Guus Wijngaards. 2014. Assessing student teachers' reflective writing through quantitative content analysis. *European Journal of Teacher Education* 37, Nr. 3 (3. Juli): 348–373. doi:10.1080/02619768.2014.892479.

- Polyzos, Serafeim und Dionysios Minetos. 2011. An ordinal regression analysis of tourism enterprises' location decisions in Greece. *Anatolia* 22, Nr. 1 (April): 102–119. doi:10.1080/13032917.2011.556225.
- Prilla, M., V. Pammer und B. Krogstie. 2013. Fostering Collaborative Redesign of Work Practice: Challenges for Tools Supporting Reflection at Work. In: *Proceedings of the European Conference on Computer Supported Cooperative Work (ECSCW 2013)*.
- Prilla, Michael. 2014. Collaborative Reflection Support at Work: A Socio-Technical Design Task. *ECIS 2014 Proceedings* (7. Juni). <http://aisel.aisnet.org/ecis2014/proceedings/track12/2>.
- . 2015. Supporting Collaborative Reflection at Work: A Socio-Technical Analysis. *AIS Transactions on Human-Computer Interaction* 7, Nr. 1 (1. Januar): 1–17.
- Prilla, Michael und Oliver Blunk. 2015. Reflective TEL: Augmenting Learning Tools with Reflection Support. In: *Design for Teaching and Learning in a Networked World*, hg. von Gránne Conole, Tomaž Klobučar, Christoph Rensing, Johannes Konert, und Élise Lavoué, 626–629. Lecture Notes in Computer Science 9307. Springer International Publishing. doi:10.1007/978-3-319-24258-3\_74, [http://link.springer.com/chapter/10.1007/978-3-319-24258-3\\_74](http://link.springer.com/chapter/10.1007/978-3-319-24258-3_74) (zugegriffen: 18. November 2015).
- Prilla, Michael, Oliver Blunk, Jenny Bimrose und Alan Brown. 2014. Reflection as support for career adaptability: A concept for reflective learning in public administration. In: *Proceedings of the 4th Workshop on Awareness and Reflection in Technology-Enhanced Learning. In conjunction with the 9th European Conference on Technology Enhanced Learning: Open Learning and Teaching in Educational Communities (ECTEL 2014)*, hg. von Milos Kravcik, Alexander Mikroyannidis, Viktoria Pammer, Michael Prilla, Thomas Daniel Ullmann, und Fridolin Wild, 65–71. Graz, Austria, 15. September.
- Prilla, Michael, Martin Degeling und Thomas Herrmann. 2012. Collaborative Reflection at Work: Supporting Informal Learning at a Healthcare Workplace. In: *Proceedings of the 17th ACM International Conference on Supporting Group Work*, 55–64. GROUP '12. New York, NY, USA: ACM. doi:10.1145/2389176.2389185, <http://doi.acm.org/10.1145/2389176.2389185> (zugegriffen: 11. Juni 2019).
- Prilla, Michael und Thomas Herrmann. 2017. Challenges for Socio-technical Design in Health Care: Lessons Learned From Designing Reflection Support. In: *Designing Healthcare That Works: A Sociotechnical Approach*, hg. von Mark Ackerman, Sean Goggins, Thomas Herrmann, Michael Prilla, und Christian Stary, 149–166. Elsevier.
- Prilla, Michael, Thomas Herrmann und Martin Degeling. 2013. Collaborative Reflection for Learning at the Healthcare Workplace. In: *Computer-Supported Collaborative Learning at the Workplace: CSCL@Work*, hg. von Sean P. Goggins, Isa Jahnke, und Volker Wulf, 139–165. Computer-Supported Collaborative Learning Series. Boston, MA: Springer US. doi:10.1007/978-1-4614-1740-8\_7, [https://doi.org/10.1007/978-1-4614-1740-8\\_7](https://doi.org/10.1007/978-1-4614-1740-8_7) (zugegriffen: 11. Juni 2019).
- Prilla, Michael, Kristin Knipfer, Martin Degeling, Ulrike Cress und Thomas Herrmann. 2011. Computer Support for Collaborative Reflection on Captured Teamwork Data. In: , Vol-790:56–61. CEUR-WS.
- Prilla, Michael, Alexander Nolte, Oliver Blunk, Dennis Liedtke und Bettina Renner. 2015. Analyzing Collaborative Reflection Support: A Content Analysis Approach. In: *ECSCW 2015: Proceedings of the 14th European Conference on*

- Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*, hg. von Nina Boulus-Rødje, Gunnar Ellingsen, Tone Bratteteig, Margunn Aanestad, und Pernille Bjørn, 123–142. Springer International Publishing. doi:10.1007/978-3-319-20499-4\_7, [http://link.springer.com/chapter/10.1007/978-3-319-20499-4\\_7](http://link.springer.com/chapter/10.1007/978-3-319-20499-4_7) (zugegriffen: 14. Dezember 2015).
- Prilla, Michael und Bettina Renner. 2014. Supporting Collaborative Reflection at Work: A Comparative Case Analysis. In: *Proceedings of ACM Conference on Group Work (GROUP 2014)*. ACM.
- Puussaar, Aare, Adrian K. Clear und Peter Wright. 2017. Enhancing Personal Informatics Through Social Sensemaking. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6936–6942. CHI '17. New York, NY, USA: ACM. doi:10.1145/3025453.3025804, <http://doi.acm.org/10.1145/3025453.3025804> (zugegriffen: 29. Mai 2019).
- Raelin, Joseph A. 2002. I don't have time to think!" versus the art of reflective practice. *Reflections* 4, Nr. 1: 66–79.
- Reid, D. J., J. Zhang und Q. Chen. 2003. Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning* 19, Nr. 1: 9–20. doi:10.1046/j.0266-4909.2003.00002.x.
- Renner, Bettina, Michael Prilla, Ulrike Cress und Joachim Kimmerle. 2016. Effects of prompting in reflective learning tools: Findings from experimental field, lab, and online studies. *Frontiers in Psychology* 7: 820.
- Renner, Bettina, Gudrun Wesiak, Viktoria Pammer-Schindler, Michael Prilla, Lars Müller, Dalia Morosini, Simone Mora, Nils Faltin und Ulrike Cress. 2019. Computer-supported reflective learning: how apps can foster reflection at work. *Behaviour & Information Technology* 0, Nr. 0 (20. März): 1–21. doi:10.1080/0144929X.2019.1595726.
- Reynolds, Michael. 1999. Grasping the Nettle: Possibilities and Pitfalls of a Critical Management Pedagogy. *British Journal of Management* 10, Nr. 2: 171–184. doi:10.1111/1467-8551.00118.
- Rink, Elizabeth, Ray Tricker und S. Marie Harvey. 2007. Onset of Sexual Intercourse among Female Adolescents: The Influence of Perceptions, Depression, and Ecological Factors. *Journal of Adolescent Health* 41, Nr. 4 (1. Oktober): 398–406. doi:10.1016/j.jadohealth.2007.04.017.
- Roberts, Andrew. 2009. Encouraging reflective practice in periods of professional workplace experience: the development of a conceptual model. *Reflective Practice* 10, Nr. 5 (1. November): 633–644. doi:10.1080/14623940903290703.
- Rosenshine, Barak, Carla Meister und Saul Chapman. 1996. Teaching Students to Generate Questions: A Review of the Intervention Studies. *Review of Educational Research* 66, Nr. 2 (1. Juni): 181–221. doi:10.3102/00346543066002181.
- Ross, Joel, Lilly Irani, M. Six Silberman, Andrew Zaldivar und Bill Tomlinson. 2010. Who Are the Crowdworkers?: Shifting Demographics in Mechanical Turk. In: *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, 2863–2872. CHI EA '10. New York, NY, USA: ACM. doi:10.1145/1753846.1753873, <http://doi.acm.org/10.1145/1753846.1753873> (zugegriffen: 11. November 2015).
- Ryan, C. Anthony. 2010. Reflective Inquiry in the Medical Profession. In: *Handbook of Reflection and Reflective Inquiry: Mapping a Way of Knowing for Professional Reflective Inquiry*, hg. von Nona Lyons, 101–130. Boston, MA: Springer US. doi:10.1007/978-0-387-85744-2\_6, [https://doi.org/10.1007/978-0-387-85744-2\\_6](https://doi.org/10.1007/978-0-387-85744-2_6) (zugegriffen: 16. Juni 2019).

- Saksono, Herman und Andrea G. Parker. 2017. Reflective Informatics Through Family Storytelling: Self-discovering Physical Activity Predictors. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5232–5244. CHI '17. New York, NY, USA: ACM. doi:10.1145/3025453.3025651, <http://doi.acm.org/10.1145/3025453.3025651> (zugegriffen: 1. Dezember 2017).
- Sandars, John. 2009. The use of reflection in medical education: AMEE Guide No. 44. *Medical Teacher* 31, Nr. 8 (1. Januar): 685–695. doi:10.1080/01421590903050374.
- Santanen, Eric L. 2005. Resolving Ideation Paradoxes: Seeing Apples as Oranges through the Clarity of ThinkLets. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences, 2005. HICSS '05*, 16c–16c. Januar. doi:10.1109/HICSS.2005.521.
- Santanen, Eric L., R.O. Briggs und G. de Vreede. 2000. The cognitive network model of creativity: a new causal model of creativity and a new brainstorming technique. In: *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000*, 10 pp. vol.1-. Januar. doi:10.1109/HICSS.2000.926895.
- Savage, Neil. 2012. Gaining Wisdom from Crowds. *Commun. ACM* 55, Nr. 3 (März): 13–15. doi:10.1145/2093548.2093553.
- Scardamalia, Marlene, Carl Bereiter und Rosanne Steinbach. 1984. Teachability of Reflective Processes in Written Composition. *Cognitive Science* 8, Nr. 2 (1. April): 173–190. doi:10.1207/s15516709cog0802\_4.
- Schoenbelen, Tyler und Victor Kuperman. 2010. Using Amazon Mechanical Turk for linguistic research. *Psihologija* 43, Nr. 4: 441–464.
- Schön, Donald A. 1983. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books.
- Schwarz, Norbert und Daphna Oyserman. 2001. Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. *American Journal of Evaluation* 22, Nr. 2 (1. Juni): 127–160. doi:10.1177/109821400102200202.
- Scott, Susanne G. 2010. Enhancing Reflection Skills Through Learning Portfolios: An Empirical Test. *Journal of Management Education* 34, Nr. 3: 430–457.
- Shah, Jami J., Steve M. Smith und Noe Vargas-Hernandez. 2003. Metrics for measuring ideation effectiveness. *Design Studies* 24, Nr. 2 (1. März): 111–134. doi:10.1016/S0142-694X(02)00034-0.
- Shields, E. 1995. Reflection and learning in student nurses. *Nurse Education Today* 15, Nr. 6 (1. Dezember): 452–458. doi:10.1016/S0260-6917(95)80058-1.
- Sinclair, Kit und Harrison Tse. 2008. Writing Reflective Journals. In: *Reflective Teaching & Learning in the Health Professions*, 84–101. John Wiley & Sons, Ltd. doi:10.1002/9780470690550.ch6, <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470690550.ch6> (zugegriffen: 13. Juni 2019).
- Skinner, B. F. 1957. *Verbal behavior*. Verbal behavior. East Norwalk, CT, US: Appleton-Century-Crofts. doi:10.1037/11256-000.
- Slovák, Petr, Christopher Frauenberger und Geraldine Fitzpatrick. 2017. Reflective Practicum: A Framework of Sensitising Concepts to Design for Transformative Reflection. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2696–2707. CHI '17. New York, NY, USA: ACM. doi:10.1145/3025453.3025516, <http://doi.acm.org/10.1145/3025453.3025516> (zugegriffen: 29. Mai 2019).
- Spais, George S und Konstantinos Z Vasileiou. 2006. An ordinal regression analysis for the explanation of consumer overall satisfaction in the food-marketing context:

- The managerial implications to consumer strategy management at a store level. *Journal of Database Marketing & Customer Strategy Management* 14, Nr. 1 (1. Oktober): 51–73. doi:10.1057/palgrave.dbm.3250036.
- Stichter, Janine P., Timothy J. Lewis, Tiffany A. Whittaker, Mary Richter, Nanci W. Johnson und Robert P. Trussell. 2009. Assessing Teacher Use of Opportunities to Respond and Effective Classroom Management Strategies Comparisons Among High- and Low-Risk Elementary Schools. *Journal of Positive Behavior Interventions* 11, Nr. 2 (1. April): 68–81. doi:10.1177/1098300708326597.
- Stoszkowski, John und Dave Collins. 2017. Using shared online blogs to structure and support informal coach learning—part 1: a tool to promote reflection and communities of practice. *Sport, Education and Society* 22, Nr. 2 (17. Februar): 247–270. doi:10.1080/13573322.2015.1019447.
- Sukhov, Alexandre. 2018. The role of perceived comprehension in idea evaluation. *Creativity and Innovation Management* 27, Nr. 2: 183–195. doi:10.1111/caim.12262.
- Suri, Siddharth und Duncan J. Watts. 2011. Cooperation and Contagion in Web-Based, Networked Public Goods Experiments. *PLOS ONE* 6, Nr. 3 (11. März): e16836. doi:10.1371/journal.pone.0016836.
- Swan, Elaine. 2008. Let's not get too personal: critical reflection, reflexivity and the confessional turn. *Journal of European Industrial Training* 32, Nr. 5 (6. Juni): 385–399. doi:10.1108/03090590810877102.
- Tang, Catherine. 2000. Reflective Diaries as A means of Facilitating and Assessing Reflection. *Pacific Rim Conference on Higher Education Planning and Assessment, Hilo, Hawaii*, 2 (3. Juni).
- Thaler, Richard H. und Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.
- Thayer-Bacon, Barbara J. 1993. Caring and Its Relationship to Critical Thinking. *Educational Theory* 43, Nr. 3: 323–340. doi:10.1111/j.1741-5446.1993.00323.x.
- Thillmann, Hubertina, Josef Künsting, Joachim Wirth und Detlev Leutner. 2009. Is it Merely a Question of “What” to Prompt or Also “When” to Prompt?: The Role of Point of Presentation Time of Prompts in Self-Regulated Learning. *Zeitschrift für Pädagogische Psychologie* 23, Nr. 2 (1. Januar): 105–115. doi:10.1024/1010-0652.23.2.105.
- Thomas, Lisa, Pam Briggs, Finola Kerrigan und Andrew Hart. 2018. Exploring digital remediation in support of personal reflection. *International Journal of Human-Computer Studies* 110 (1. Februar): 53–62. doi:10.1016/j.ijhcs.2017.10.002.
- Thompson, Mark. 2005. Structural and Epistemic Parameters in Communities of Practice. *Organization Science* 16, Nr. 2: 151–164.
- Tigelaar, Dineke E. H., Diana H. J. M. Dolmans, Paulien C. Meijer, Willem S. de Grave und Cees P. M. van der Vleuten. 2008. Teachers' Interactions and their Collaborative Reflection Processes during Peer Meetings. *Advances in Health Sciences Education* 13, Nr. 3 (1. August): 289–308. doi:10.1007/s10459-006-9040-4.
- Titi Amayah, Angela. 2013. Determinants of knowledge sharing in a public sector organization. *Journal of Knowledge Management* 17, Nr. 3 (24. Mai): 454–471. doi:10.1108/JKM-11-2012-0369.
- Ullmann, Thomas Daniel. 2015a. Automated detection of reflection in texts. A machine learning based approach. Phd, The Open University, April. <http://oro.open.ac.uk/45402/> (zugegriffen: 7. März 2016).

- . 2015b. Keywords of written reflection - a comparison between reflective and descriptive datasets. In: *CEUR Workshop Proceedings*, hg. von Milos Kravcik, Alexander Mikroyannidis, Viktoria Pammer, Michael Prilla, und Thomas Daniel Ullmann, 1465:83–96. Toledo, Spain, 9. Oktober. <http://ceur-ws.org/Vol-1465/paper8.pdf> (zugegriffen: 1. März 2016).
- Vakharia, Donna und Matthew Lease. 2013. Beyond AMT: An Analysis of Crowd Work Platforms. *arXiv:1310.1672 [cs]* (7. Oktober). <http://arxiv.org/abs/1310.1672> (zugegriffen: 12. November 2015).
- Verplanken, Bas, Oddgeir Friberg, Catharina E. Wang, David Trafimow und Kristin Woolf. 2007. Mental habits: Metacognitive reflection on negative self-thinking. *Journal of Personality and Social Psychology* 92, Nr. 3: 526–541. doi:10.1037/0022-3514.92.3.526.
- Vince, Russ. 2002. Organizing Reflection. *Management Learning* 33, Nr. 1 (1. März): 63–78. doi:10.1177/1350507602331003.
- Vivacqua, Adriana S. und Marcos R. S. Borges. 2012. Taking advantage of collective knowledge in emergency response systems. *Journal of Network and Computer Applications* 35, Nr. 1. Collaborative Computing and Applications (1. Januar): 189–198. doi:10.1016/j.jnca.2011.03.002.
- Vogt, Eric E, Juanita Brown und David Isaacs. 2003. *The Art of Powerful Questions: Catalyzing, Insight, Innovation, and Action*. Whole Systems Associates.
- Wagner, C. und P. Prasarnphanich. 2007. Innovating Collaborative Content Creation: The Role of Altruism and Wiki Technology. In: *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, 18–18. Januar. doi:10.1109/HICSS.2007.277.
- Walker, David. 1985. Writing and Reflection. In: *Reflection: Turning Experience Into Learning*, hg. von David Boud, Rosemary Keogh, und David Walker, 52–68. Routledge.
- Wallman, Andy, Asa Kettis Lindblad, Stina Hall, Annika Lundmark und Lena Ring. 2008. A categorization scheme for assessing pharmacy students' levels of reflection during internships. *American Journal of Pharmaceutical Education* 72, Nr. 1: 05.
- Walsh, Anita. 2009. Modes of Reflection: Is it possible to use both individual and collective reflection to reconcile the 'three-party knowledge interests' in workplace learning? *European Journal of Education* 44, Nr. 3 (1. September): 385–398. doi:10.1111/j.1465-3435.2009.01389.x.
- Wang, Kai, Hui Wang und Yu Tao. 2017. Combining Ideas in Crowdsourced Idea Generation. *Foundations of Management* 9, Nr. 1 (23. Februar): 203–212. doi:10.1515/fman-2017-0016.
- Ward, Edward A. 1993. Generalizability of Psychological Research from Undergraduates to Employed Adults. *The Journal of Social Psychology* 133, Nr. 4 (1. August): 513–519. doi:10.1080/00224545.1993.9712176.
- Wasko, Molly McLure und Samer Faraj. 2005. Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly* 29, Nr. 1: 35–57.
- Webb, Oliver J. und Frank F. Eves. 2007. Promoting stair climbing: effects of message specificity and validation. *Health Education Research* 22, Nr. 1 (1. Februar): 49–57. doi:10.1093/her/cyl045.
- Weick, Karl E. 1995. *Sensemaking in Organizations*. SAGE.

- Weinmann, Markus, Christoph Schneider und Jan vom Brocke. 2016. Digital Nudging. *Business & Information Systems Engineering* 58, Nr. 6 (1. Dezember): 433–436. doi:10.1007/s12599-016-0453-1.
- Wenger, E., R.A. McDermott und W. Snyder. 2002. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. NetLibrary Inc. Harvard Business School Press. <https://books.google.de/books?id=m1xZuNq9RygC>.
- Wenger, Etienne. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.
- West, Michael A. 1996. Reflexivity and Work Group Effectiveness: a Conceptual Integration. In: *Handbook of Work Group Psychology*, 555–579. Chichester: Wiley.
- Wever, B. De, T. Schellens, M. Valcke und H. Van Keer. 2006. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education* 46, Nr. 1: 6–28. doi:<http://dx.doi.org/10.1016/j.compedu.2005.04.005>.
- Winchester, Tiffany M. und Maxwell Winchester. 2011. Exploring the impact of faculty reflection on weekly student evaluations of teaching. *International Journal for Academic Development* 16, Nr. 2 (1. Juni): 119–131. doi:10.1080/1360144X.2011.568679.
- Wise, Alyssa Friend, Yuting Zhao und Simone Nicole Hausknecht. 2013. Learning Analytics for Online Discussions: A Pedagogical Model for Intervention with Embedded and Extracted Analytics. In: *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 48–56. LAK '13. New York, NY, USA: ACM. doi:10.1145/2460296.2460308, <http://doi.acm.org/10.1145/2460296.2460308> (zugegriffen: 25. Juni 2015).
- Woerkom, Marianne van und Marcel Croon. 2008. Operationalising critically reflective work behaviour. *Personnel Review* 37, Nr. 3 (11. April): 317–331. doi:10.1108/00483480810862297.
- Wong, Bernice Y. L. 1985. Self-Questioning Instructional Research: A Review. *Review of Educational Research* 55, Nr. 2 (1. Juni): 227–268. doi:10.3102/00346543055002227.
- Wopereis, Iwan G. J. H., Peter B. Sloep und Sybilla H. Poortman. 2010. Weblogs as instruments for reflection on action in teacher education. *Interactive Learning Environments* 18, Nr. 3 (1. September): 245–261. doi:10.1080/10494820.2010.500530.
- Xie, Luning und Pengzhu Zhang. 2010. Idea Management System for Team Creation. *ISSN 1796-217X* 5, Nr. 11. Journal of Software: 1187.
- Yang, Shih-Hsien. 2009. Using Blogs to Enhance Critical Reflection and Community of Practice. *Journal of Educational Technology & Society* 12, Nr. 2: 11–21.
- Ye, Jonathan, Ivo Blohm, Ulrich Breschneider, Suparna Goswami, Jan Marco Leimeister und Helmut Krcmar. 2016. Promoting the Quality of User Generated Ideas in Online Innovation Communities: A Knowledge Collaboration Perspective. In: *Proceedings of the International Conference on Information Systems - Digital Innovation at the Crossroads, ICIS 2016*. AIS Electronic Library (AISel): Association for Information Systems, 11. Dezember. <http://aisel.aisnet.org/icis2016/SocialMedia/Presentations/2/> (zugegriffen: 16. Mai 2019).
- Yoo, Daisy, Alina Hultgren, Jill Palzkill Woelfer, David G. Hendry und Batya Friedman. 2013. A Value Sensitive Action-reflection Model: Evolving a Co-design Space with Stakeholder and Designer Prompts. In: *Proceedings of the*

- SIGCHI Conference on Human Factors in Computing Systems*, 419–428. CHI '13. New York, NY, USA: ACM. doi:10.1145/2470654.2470715, <http://doi.acm.org/10.1145/2470654.2470715> (zugegriffen: 5. Februar 2016).
- Zimmerman, Barry J. 1990. Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist* 25, Nr. 1 (1. Januar): 3–17. doi:10.1207/s15326985ep2501\_2.



